STATISTICAL ASSESSMENT OF SAMPLING FREQUENCY REQUIREMENTS FOR SELECTED ASPECTS OF THE MISA PROGRAM

Prepared for the
MISA Advisory Committee

November 1988

Gartner
Lee

STATISTICAL ASSESSMENT OF SAMPLING FREQUENCY REQUIREMENTS FOR SELECTED ASPECTS OF THE MISA PROGRAM VOLUME 1

PREPARED FOR:
MISA ADVISORY COMMITTEE

PREPARED BY:
GARTNER LEE LIMITED

GLL 88-194

NOVEMBER, 1988

DISTRIBUTION:
CC: 1 CLIENT
3 FILE

STATISTICAL ASSESSMENT OF SAMPLING FREQUENCY REQUIREMENTS FOR SELECTED ASPECTS OF THE MISA PROGRAM VOLUME 1

PREPARED FOR:
MISA ADVISORY COMMITTEE

PREPARED BY:
GARTNER LEE LIMITED

GLL 88-194

NOVEMBER, 1988

DISTRIBUTION:
CC: 1 CLIENT
3 FILE

Gartner
Lee

140 Renfrew Drive,
Suite 102
Markham, Ontario
L3R 8B6
Fax (416) 477-1456
Telex 06-986278

(416) 477-8400

Professional Services in Environmental Management

- *Environmental Strategies*
- *Resources Planning*
- *Waste Management*
- *Hydrogeology*
- *Biology*
- *Water Quality*
- *Engineering Geology*
- *Spill Response*

November 21, 1988 GLL 88-194

MISA Advisory Committee
Suite 502
112 St. Clair Ave. W.
Toronto, Ontario

Attention: Mr. Doug Vallery
Executive Co-ordinator

Dear Sirs:

Re: Statistical Assessment of Sampling Frequency Requirements for Selected Aspects of the MISA Program

We are pleased to provide our final report (volume 1) and appendix (volume 2) addressing statistical questions about the MISA monitoring programs. The report demonstrates the importance of sampling frequency in improving accuracy and confidence interval width and determining presence/absence of compounds in effluents.

Yours very truly,

GARTNER LEE LIMITED

J.E. O'Neill, B.Sc.
Hydrologist,
Senior Consultant

JEO:tmc
Encl.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

APPENDICES

(Contained in a separately bound volume)

EXECUTIVE SUMMARY

Selected components of the MISA program relating to sampling frequency requirements were investigated, specifically:

- estimation of monthly means for use in the development of BATEA effluent limits, and
- data characterization for the determination of presence/absence of a compound.

Basic statistical procedures and approaches used by various researchers were also employed here to investigate the above aspects of the MISA program.

The first study component involved investigation of the effect of four relative levels of variability of an industrial effluent constituent on the resulting accuracy of estimates of the mean. A package of computer programs was developed to assist in the analysis. The problem of a lack of actual data for use in our analysis was overcome by using a series of artificially generated data sets to represent different levels of variability.

The second study component involved the determination of the minimum sampling frequency capable of detecting various levels of pollutant occurrence. Basic statistical procedures and computer programs were used to generate artificial data and repeatedly sample the data bases to estimate sampling efficiencies.

In the case of industrial effluent constituents with high and very high variability (i.e., coefficient of variation (CV) >60%), the study found that greater than thrice weekly sampling was required to meet the assumed accuracy goal of (±25%) for estimation of monthly means to be used for development of BATEA effluent limits. At least thrice weekly sampling was required for medium variability (30% < CV ≤ 60%) and at least weekly for low variability (CV ≤30%).

The minimum sampling frequencies required to identify the presence of constituents (at least 80%* of the time) for various values of θ (probability of a constituent being above the detection limit on any given day) were identified as follows:

Frequency	Suitable For θ
Monthly	>.15
Bi-monthly	>.25
Quarterly	>.35
Semi-annual	>.50 (*at 75%) or >.40 (*at 64%)

The methods presented in this study can be used to assess actual industrial monitoring data as they become available. The MISA objectives should be reviewed in terms of the statistical requirements to achieve stated goals. Specifically, the accuracy and precision requirements for BATEA data sets must be determined (or values assumed). Also the minimum θ that the characterization program is to detect should be specified along with the minimum probability for detection that is acceptable. As these values are refined, the calculations in this report should be updated.

1.0 INTRODUCTION

1.1 BACKGROUND

The MISA Advisory Committee (MAC) is a group of independent technical and environmental experts. MAC was established in November, 1986 to review draft regulations and to provide advice and recommendations to the Minister of the Environment concerning the Municipal/Industrial Strategy for Abatement (MISA) program.

MAC is concerned about the statistical validity of certain sampling frequencies proposed in the draft regulations. The Committee wishes to determine monitoring criteria necessary to develop a data base which will provide for the achievement of MISA program objectives.

Gartner Lee Limited was retained to conduct a statistical assessment of selected components of the MISA monitoring approach.

1.2 OBJECTIVES

The three objectives for this study were:

1. to specify statistically justifiable sampling frequencies and protocols required to provide valid data (concentrations and loadings of conventional and organic parameters) on which to base *BATEA**,

2. to specify the data base characteristics (i.e. data quantity, quality and frequency) necessary to determine, within reasonable *confidence limits*, the presence or absence of all Effluent Monitoring Priority Pollutant List (EMPPL) compounds in effluents, and

3. to specify the data base characteristics necessary to identify compounds in effluents which are not on the EMPPL.

---

*Terms in italics are defined in the glossary contained at the end of this report

1.3 APPROACH

The questions raised by the MISA Advisory Committee deal with design of water quality monitoring programs, specifically sampling frequency. This topic has received considerable attention in recent years (e.g., Ward, et al, 1979; Loftis, et al, 1983; Loftis, et al, 1987). The following general approach (or some variation of it) is commonly used in designing monitoring networks:

1. Define monitoring objectives,
2. Express objectives in statistical terms,
3. Determine parameters, frequency, station locations, etc. to achieve program objectives,
4. Implement monitoring program,
5. Report,
6. Evaluate and adjust network at regular intervals.

In this study, we are primarily concerned with item 3 relating to the specification of sampling frequency. To examine this aspect; however, it is also necessary to look at item 2, namely, expressing certain MISA objectives or sub-objectives in statistical terms.

To provide a definitive specification of sampling frequency in a program, it is necessary to have a suitable set of data (i.e. a preliminary data set) that will provide the necessary statistical information (e.g. water quality variability). Usually no data or only limited data are available at the beginning of a study thus data collected elsewhere or statistical "rules of thumb" are often used to design a preliminary monitoring phase which is then re-designed once suitable data (usually one year) have been collected.

The preliminary data set must be sufficiently detailed to allow the use of statistical techniques to characterize the data in terms such as *mean*, *standard deviation*, confidence limits and *significance level.* Where water quality is variable due to random and/or systematic variability, statistical values (such as the mean), are estimates of the true values.

The difference between estimated and true values can be calculated statistically. This difference is related to the variability of the water quality parameter being measured, which in turn is related to:

- effluent quality and quantity related to the process,
- sampling, preservation and handling procedures,
- laboratory analytical techniques, and
- quality assurance and protocols.

When the preliminary data set has been obtained it is used to determine the inherent variability of the data. Once this variability is known the sampling frequency can be modified to achieve stated precision goals for the program.

In the present study, only limited data were available concerning water quality in industrial effluents (i.e., the preliminary data set is not sufficient). This limitation was addressed by simulating a data base considered to be representative of general classes of industries. The simulated data base was then sampled according to various scenarios (e.g. weekly, thrice weekly) to assess effectiveness of each.

1.4 STUDY SCOPE AND ASSUMPTIONS

The scope of this study was limited to specific questions of sampling frequency. Other issues such as station location, sample collection and lab analysis, although important considerations in monitoring design, were not addressed in this report. The question of how the information generated by the monitoring programs will be used was not addressed. For example, it was possible to quantify *confidence intervals* and *accuracy* for various sampling frequencies and parameter variabilities. However, it was not possible to assess whether a particular confidence interval and *confidence level* or accuracy will be suitable for development of (BATEA) effluent limits. To overcome this limitation, assumptions were made concerning the required accuracy and *precision* for development of BATEA effluent limits.

1.5 REPORT ORGANIZATION

The report is presented in two separately bound volumes. The main report is contained in Volume 1 (this volume). Volume 2 contains the technical appendices.

The main report contains an Executive Summary and Glossary of Terms to facilitate understanding of the report. The technical appendices are comprised of three parts. An overview assessment of the expected variability of industrial effluent quality and quantity is presented in Appendix A. The documentation of the computer "model" developed for this project is contained in Appendix B. Executable files and sample data for the *BASIC* programs and *MINITAB macros* are included on a floppy disk which accompanies Volume 2 of this report. Appendix C contains the simulated data bases used as examples in the report.

2.0 METHODS

2.1 INDUSTRIAL EFFLUENTS

An overview assessment was undertaken by Zenon Environmental Inc. (Canning, 1988) to provide insight and background information concerning the range of *variability* likely to be encountered in industrial effluent quality and quantity. The assessment was limited in scope, intended to provide an approximate range of expected conditions which could be used as a framework for investigating sample frequency requirements.

The assessment was accomplished by a review of selected references. The results are presented in Appendix A.

2.2 STATISTICS

The procedures used to investigate sample frequency requirements from a network design perspective were based primarily on the work of several investigators at Colorado State University. These network design procedures have been presented in several references, e.g. (Ward, et al, 1979; Sanders, et al, 1979). The statistical formulae, theorems, etc. used in calculation (e.g. confidence intervals, means, etc.) are available in any statistical text. For this project Freund, 1962, Spiegel, 1961; and Yevjevich, 1972 were used as general references for statistics formulae.

2.3 COMPUTER MODEL

A computer model was developed to accomplish the following:

1. simulate a data base,
2. sample the data base in different ways, and
3. calculate descriptive statistics for various sampling scenarios.

The general purpose of the model was to demonstrate the statistical principles involved in determining sample frequency requirements. In the absence of an adequate preliminary data set now, it was decided to include a simulation (generation of artificial data) option. This component generates an artificial data base of predetermined characteristics.

As real data become available, they can be analyzed by these programs and the results used to refine the monitoring program.

Two complementary sets of programs were developed, one programmed in BASIC and the other programmed in MINITAB (a statistical analysis package). The programs were designed to work together, i.e. the output of one can be analyzed by the other.

BASIC was selected because it facilitated the simulation of data in a graphical format. MINITAB was selected because of its programming ability and availability of a wide range of statistical procedures[1].

1Note: There are several excellent statistical analysis packages commercially available. The use of MINITAB in this study should not be considered an endorsement.

3.0 FINDINGS

3.1 INTRODUCTION

This chapter presents the study findings relating to:

- estimation of mean monthly concentrations and loads and,
- determination of the presence or absence of compounds.

3.2 ESTIMATES OF MONTHLY MEANS

This section examines the first study objective, namely; the question of frequency of sampling necessary to provide data on which to base BATEA effluent limits. Of specific interest is the question of whether or not thrice weekly sampling can be reduced to weekly sampling while maintaining an acceptable level of accuracy in the data.

To answer this question, it is necessary to know how BATEA effluent limits will be derived and to understand the nature and variability of effluent flows and quality. Unfortunately, this information is only partly known at this time. Consequently, to complete the investigation it is necessary to make the following assumptions.

In regards to the derivation of BATEA effluent limits it has been assumed that:

1. the *relative error in estimating the mean* should not exceed ±25% (as a measure of accuracy),

2. the *relative precision in estimating the mean* should not exceed ±50%, and

3. a 95% level of confidence should be used.

(Assumption #1 and #3 are consistent with those expressed by A. Sharma (1988).

An overview assessment was undertaken by Zenon Environmental Inc. (Canning, 1988) to provide insight and background information on the range of effluent quality and quantity likely to be found in Ontario's industrial effluents (Appendix A). Based on the

overview the following relative variability levels have been assumed. The levels are expressed in terms of the *coefficient of variation* (CV).

Relative Variability

Low	$CV \leq 30\%$
Medium	$30\% < CV \leq 60\%$
High	$60\% < CV \leq 200\%$
Very High	$CV > 200\%$

3.2.1 Accuracy

Accuracy refers to the degree to which the estimates from a measurement technique agree with the true value. The value being estimated in this case is the mean monthly concentration. A measure of accuracy is provided by the following expression.

$$\Delta = \frac{\mu - \bar{x}}{\mu} * 100$$

where: Δ (delta) is the relative error, expressed as a percentage.

μ is the true mean,

$\bar{x}$ is the estimated or sample mean derived by some sampling scheme.

As stated in Section 3.2 our assumed goal is to achieve a value for Δ of less than $\pm 25\%$. Thrice weekly and weekly sampling schemes were investigated for four "typical" industries which represent a broad range of relative variability levels as explained below. A generic profile of the "typical" industry characteristics follows. Data for each industry were simulated using the programs described in Appendix B. The results appear in Appendix C3.

Industry #1 represents "Low" relative variability. Continuous operation (24 hr/day and 7 days/week) plus a high level of effluent treatment (including biological treatment) tends to produce effluents which vary over a relatively narrow range of quality and quantity compared to other groups. An example industry might be found in the petroleum refining sector.

Industry #2 represents "Medium" relative variability. This industry uses large, continuous production facilities and involves a high usage of industrial chemicals. Effluent quality and quantity exhibit relatively moderate changes; however, plant upsets can cause occasional high values. An example industry could be an inorganic chemical manufacturing plant.

Industry #3 represents "High" relative variability. This industry uses large batch processes to manufacture specialty chemicals. The process involves a large product mix, e.g. pharmaceuticals, paints, dyes, inks, etc. Minimum treatment will enhance the variability of effluent quantity and quality. An example of this industrial type may be found in the organic chemical manufacturing sector.

Industry #4 represents "very high" relative variability. Although the quality of process effluent from this industry may be relatively constant, there are large contributing areas where contaminated surface runoff can be discharged in response to rainfall and/or melt conditions.

The data for this industry were simulated using the programs in Appendix B to represent an actual industrial example. The model for this example is a base metal mining industry in the Elliot Lake area. The monitoring point includes runoff from tailings areas. The observed mean suspended solids concentration based on 83 weekly average values was 3.75 mg/L with a standard deviation of 16.34 mg/L. The coefficient of variation for this parameter was 435%.

The mean of the constituent to be modelled was 4.16 mg/L for simulated data base with a standard deviation of 13.88 mg/L. The coefficient of variation was 334%. Simulated data for this industry are contained in Appendix C3.

The "SIMULATE" program (in Appendix B) was used to sample the four (simulated) industrial data bases for thrice weekly and weekly sampling scenarios. The program also calculated the mean based on daily sampling which was assumed to be the true value. Each industry was sampled twelve times (i.e. monthly).

The results were entered into a LOTUS 1-2-3 spreadsheet for further analysis (Table 1). The estimated means were plotted against the true mean to provide a visual representation of the accuracy of the two sampling schemes. The results are shown in Figures 1a and 1b. The 1:1 line as well as the $\pm$25% error lines are shown. Values falling outside of the envelope failed to meet the assumed accuracy criterion of $\pm$25% of the true mean.

Figure 1b shows the lower concentration range for Figure 1a. Sampling scheme efficiency is summarized in Table 2. At low relative variability both sampling schemes produced results that were completely within our stated accuracy goals. At medium relative variability thrice weekly sampling still produced results which were accurate 100% of the time whereas 2 out of 12 (or 17%) of the samples from weekly sampling had unacceptable accuracy. At high relative variability both results produced some inaccuracies; however, the weekly sampling results were in error the most 8 out of 12 times or (67%). At very high relative variability both sampling schemes performed poorly.

3.2.2 Confidence Intervals

A confidence interval is defined as a range around an estimated sample statistic, such as the mean. A statement can be made about the probability of this interval to include the population mean, μ (i.e., the true mean). This statement is usually made at a 95% level of confidence. This means that on the average the true mean will be within the stated confidence interval 95 out of 100 times. Ideally, we would like to have a *confidence interval width* as close to zero as possible, i.e. we would like to be as precise as possible.

FIGURE Ia

ACCURACY OF MEAN MONTHLY ESTIMATES

THRICE WEEKLY VS WEEKLY SAMPLING

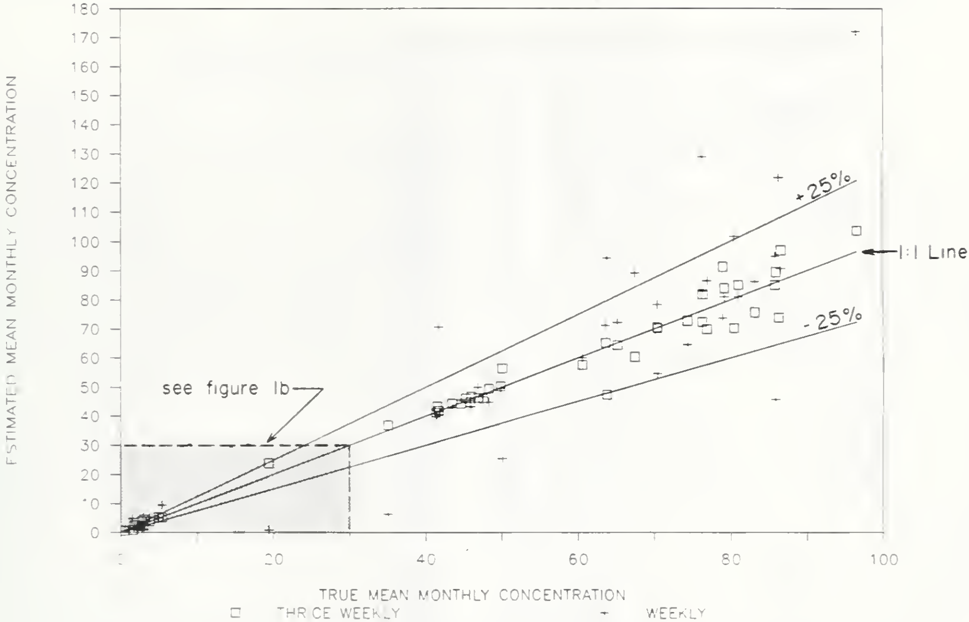


FIGURE Ib

ACCURACY OF MEAN MONTHLY ESTIMATES

THRICE WEEKLY VS WEEKLY SAMPLING

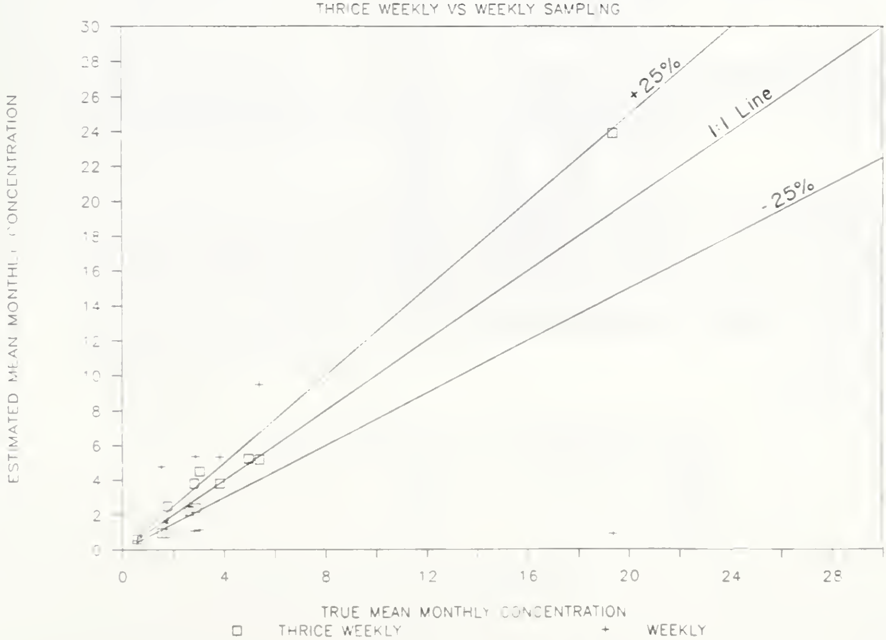


TABLE 1 COMPARISON OF TRUE MONTHLY MEAN CONCENTRATION VS. ESTIMATED

VARIABILITY	MONTH	TRUE	THRICE	WEEKLY
Low	1	41.54	43.31	40.41
	2	46.86	46.29	50.08
	3	49.85	50.40	49.64
	4	43.47	44.44	43.16
	5	44.64	44.42	45.38
	6	45.89	44.75	44.55
	7	48.32	49.41	44.90
	8	41.35	41.19	39.88
	9	47.58	46.16	46.63
	10	45.30	46.13	44.93
	11	46.01	46.82	43.19
	12	41.77	41.80	41.46
Medium	1	65.20	64.38	72.25
	2	76.30	81.85	83.00
	3	80.40	70.23	101.50
	4	67.47	60.33	89.20
	5	74.33	72.77	64.50
	6	63.63	65.23	71.25
	7	83.17	75.67	86.20
	8	85.77	85.00	94.75
	9	81.00	85.08	81.00
	10	76.87	69.92	86.50
	11	86.53	96.83	90.60
	12	79.13	83.92	81.00
High	1	41.72	41.83	70.76
	2	35.06	36.91	6.38
	3	50.09	56.41	25.39
	4	60.61	57.61	60.08
	5	96.50	103.55	172.16
	6	86.24	73.82	121.68
	7	70.41	70.24	54.65
	8	63.86	47.35	94.26
	9	85.87	89.40	45.57
	10	70.37	70.54	78.38
	11	78.98	91.17	73.71
	12	76.27	72.45	128.83
Very High	1	0.60	0.77	0.50
	2	2.90	4.38	1.00
	3	1.60	1.08	5.00
	4	19.27	23.75	0.80
	5	2.90	2.46	5.25
	6	5.33	5.15	9.25
	7	1.80	2.58	1.80
	8	2.83	3.92	1.00
	9	1.67	0.85	1.75
	10	2.60	2.23	2.50
	11	4.93	5.17	5.00
	12	3.83	3.69	5.50

TABLE 2: SAMPLING SCHEME EFFICIENCY - ACCURACY

Relative Variability Level (CV)	Coefficient of Variation (CV)	Percentage of Samples with delta >±25% Thrice	Weekly
Very High	334%	50%	58%
High	92%	8%	67%
Medium	38%	0%	17%
Low	21%	0%	0%

A relative measure of precision was obtained by expressing the confidence interval width as a percentage of the mean. In this study, we have selected the goal that the relative precision should be less than ±50% of the mean.

The confidence interval is a function of the standard deviation and the sample size Since we generally have no control over the standard deviation, it is necessary to adjust sampling size to change the width of the confidence interval.

The confidence interval width may be calculated from the confidence limits. The confidence limits are the numerical limits of the confidence interval. For samples of small size (i.e., N <30) the confidence limits about the mean ($\bar{x}$) can be calculated using student's t distribution, the standard deviation ($\hat{S}x$) and the number of samples (N) according to the expression:

$$\bar{x} - \left[t(n-1)\alpha/2 \ast \frac{\hat{S}x]}{\sqrt{N}}\right. < \bar{x} < \bar{x} + \left[t(n-1)\alpha/2 \ast \frac{\hat{S}x]}{\sqrt{N}}\right.$$

The effect of increasing sample size on the confidence interval width can be seen in a general way in Figure 2. In this graph, the effect of the standard deviation statistic has been removed by assuming it equal for all cases. The normalized confidence interval width is depicted for the 95% confidence level.

As N increases the normalized confidence interval width approaches the goal of zero. The difference between sampling thrice weekly and weekly can be seen in Figure 2. At the 95% confidence level the confidence interval width is about 2.7 times greater for weekly than for thrice weekly.

When standard deviation is considered as a factor, the difference between thrice weekly and weekly increases as standard deviation increases. This is examined in greater detail using the "typical" industrial effluent data base introduced in Section 3.2.1.

As discussed previously the "SIMULATE" program was used to generate and sample four data bases according to thrice weekly and weekly sampling schemes. The 95% confidence interval widths were calculated for each industry and sampling scheme on a monthly basis. The results appear in Appendix C3.

The relative confidence interval widths were calculated for each sampling scenario. The results are contained in Table 3. A value greater then 100% represents failure to achieve our assumed goal of confidence interval width (CIW) to be less than ±50% of the mean.

EFFECT OF SAMPLING SIZE ON NORMALIZED CONFIDENCE INTERVAL WIDTH

FIGURE 2

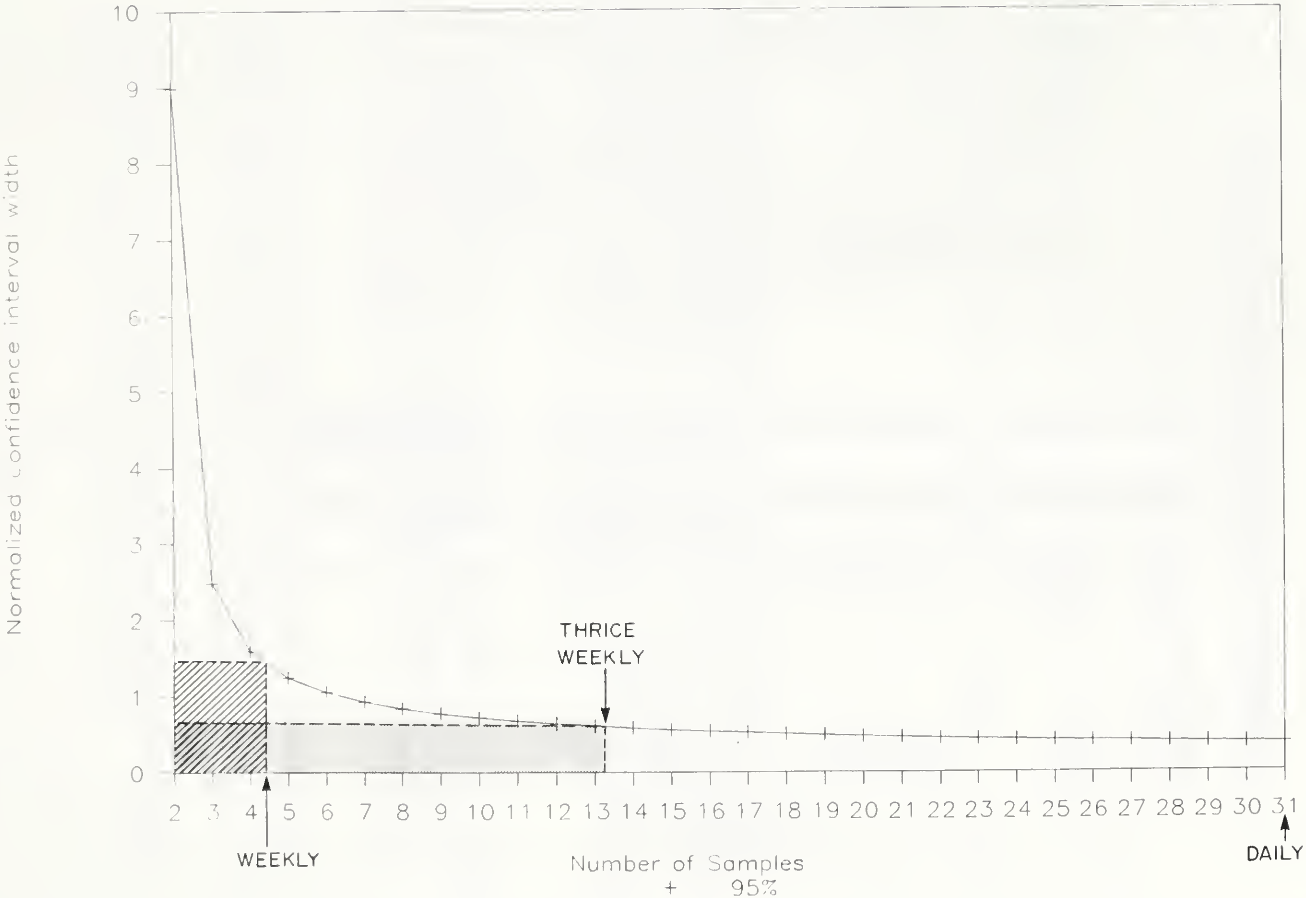


TABLE 3 RELATIVE CONFIDENCE INTERNAL WIDTH

VARIABILITY	MONTH	THRICE (%)	WEEKLY (%)
Low	1	23.6	92.0
	2	24.4	185.0
	3	33.7	66.8
	4	18.7	41.9
	5	20.9	29.7
	6	24.1	17.3
	7	15.7	43.9
	8	33.9	22.2
	9	20.1	50.5
	10	17.1	53.8
	11	27.1	22.9
	12	20.5	93.8
Medium	1	49.4	50.4
	2	44.8	231.5
	3	33.1	113.4
	4	40.5	108.8
	5	37.4	129.4
	6	41.9	92.0
	7	50.4	74.9
	8	48.0	47.0
	9	41.1	65.6
	10	25.3	45.7
	11	31.6	55.2
	12	52.7	138.6
High	1	131.8	628.0
	2	131.1	273.7
	3	123.4	231.4
	4	123.2	277.2
	5	80.3	262.3
	6	88.6	305.7
	7	132.3	249.2
	8	123.7	178.5
	9	80.0	171.6
	10	90.5	230.2
	11	108.7	156.8
	12	83.8	253.8
Very High	1	145.0	270.0
	2	297.9	22.4
	3	85.0	1447.5
	4	288.8	238.5
	5	159.3	792.4
	6	139.6	586.1
	7	264.4	237.8
	8	199.3	283.4
	9	95.2	282.0
	10	65.8	252.3
	11	145.8	331.4
	12	106.8	525.3

TABLE 4: SAMPLING SCHEME EFFICIENCY - PRECISION

Relative Confidence Interval Width	CV	Percentage of Samples with CIW >±50% of μ	
		Thrice Weekly	Weekly
Very High	334%	75%	100%
High	92%	58%	100%
Medium	38%	0%	50%
Low	21%	0%	8%

The efficiency of thrice weekly sampling versus weekly sampling is shown in Table 4. The thrice weekly sampling approach achieves our assumed goal for low and medium variable effluents whereas weekly sampling does not. At low relative variability weekly sampling fails to achieve the required precision on 1 out of 12 samples (8%). At medium variability weekly fails 6 out of 12 samples (50%). At higher variability weekly fails 100% of the time and thrice weekly fails more than 50% of the time.

3.2.3 Loading Calculation

The loading of a constituent in an industrial effluent may be defined as the rate of mass transport expressed in units of mass per time, e.g. kg/day, metric tonnes/year, etc. Loading is not measured directly, rather it is calculated using flow and concentration data.

Various methods may be used to estimate mean monthly loads. In the case of daily sampling an appropriate method would be to sum the products of the individual mean daily flows and concentrations and divide by the total by the number of samples.

When such a data base is available then the confidence intervals for the mean can be calculated in the same manner as discussed previously in Section 3.2.2:

When loading is determined by the multiplication of monthly means (flow and concentration) the reliability of the calculated value is found by applying the theory propagation of errors (Overman & Clark, 1960).

Assuming that concentration and flow are independent variables, the variance of the load (σ_L^2) is obtained by the equation:

$$\sigma_L^2 = C^2 \times \sigma_Q^2 + Q^2 \times \sigma_C^2 \qquad [1]$$

where: L is load
C is concentration,
Q is flow,
σ^2 is variance, and
$\times$ denotes multiplication.

In cases where concentration and flow are not independent variables the variance of the load is given by:

$$\sigma_L^2 = C^2 \times \sigma_Q^2 + Q^2 \times \sigma_C^2 + 2QC \times \sigma_{QC} \qquad [2]$$

where: σ_{QC} is the flow-concentration covariance and other parameters are as defined above.

Equation 2 can be simplified as follows:

$$\frac{\sigma_L^2}{C^2 Q^2} = \frac{\sigma_Q^2}{Q^2} + \frac{\sigma_C^2}{C^2} + \frac{2\sigma_{QC}}{CQ}$$

$$= CV_Q^2 + CV_C^2 + \frac{2\,\sigma_{QC}}{CQ}$$

But L = QC so $\dfrac{\sigma_L^2}{C^2Q^2} = \dfrac{\sigma_L^2}{L^2} = CV_L^{\;2}$

$$\therefore \quad CV_L^2 = CV_Q^2 + CV_C^2 + \frac{2\,\sigma_{QC}}{L} \qquad [3]$$

Where CV is the coefficient of variation.

Similarly, for the independent case, it can be shown that

$$CV_L^{\,2} = CV_Q^{\,2} + CV_C^{\,2} \qquad [4]$$

The above relationship can be illustrated by an example. Assuming that flow and concentration data are independent then [4] can be used and the mean monthly load ($\bar{L}$) is to be calculated using the mean monthly flow ($\bar{Q}$) and mean monthly concentration ($\bar{C}$) using the following values:

$\bar{C}$ = 4.2 mg/L
CV_c = 334%

$\bar{Q}$ = 13,088.9 m3/s
CV_Q = 9.3%

$\bar{L}$ = $\bar{C} \times \bar{Q}$

$$\bar{L} = \frac{4.2 \text{ mg}}{\text{L}} \times \frac{13{,}088.9 \text{ m3}}{\text{S}}$$

$$= \frac{4.2 \text{ mg}}{\text{L}} \times \frac{13{,}088.9}{\text{S}} \times 1000 \text{ L}$$

$$\bar{L} = \frac{5.5 \times 10^7 \text{ mg}}{S}$$

$$CV_L = CV_Q^2 + CV_C^2$$

$$= 9.3^2 + 334^2$$

$$CV_L = 334\%$$

If the coefficient of variation of the flow is increased to 25% (e.g., less accurate measurement techniques) and the calculations repeated then the CV_L^2 is:

$$CV_Q^2 + CV_C^2$$

$$= 25^2 + 334$$

$$CV_L = 335\%$$

In other words when the coefficient of variation is high for one element of the calculation with respect to the other then that element is predominantly responsible for the variation in the calculated product. In such cases, reducing the accuracy for the parameter with a very low CV will result in only a small increase in the CV of the calculated product.

3.2.4 Discussion

A comparison of the suitability of weekly and thrice weekly sampling for achieving the assumed accuracy and precision goals is given below.

	Accuracy	Precision
Weekly	- suitable for low variability industries	- not suitable for any level of variability

Thrice	- suitable for low and medium variability levels	- suitable for low and medium variability levels

For industries with high levels of variability, a sampling frequency greater than thrice weekly is required.

It is worthwhile to note that parameters within an industry will have different variability levels. Thus, a particular industry may very well require different sampling frequencies for each parameter or group of parameters. When designing a program with different sampling requirements, it is possible to:

a) design for the most demanding parameter, or
b) design to some acceptable middle condition.

Designing to the most demanding parameter will ensure that the required data are obtained for all cases although costs and logistics will be high and some unnecessary data will be obtained. Designing to some middle condition will reduce costs and logistics but may result in loss of some required data.

In regards to approach (a), a goal of uniform precision in the data may be applicable to the MISA program. For example, sampling at a thrice weekly frequency for the highest variability industry will produce results that are accurate to ±X%. Sampling at the same frequency at a low variability industry will produce results that are accurate to ±Y% where Y <X. If we accept that X accuracy can equal Y accuracy, then the sampling frequency requirements can be reduced at the low variability industry. The same argument would apply to precision.

Finally, in the case of loading estimates, there is the potential for reducing the required accuracy of flow measurement devices where the CV for the constituent of concern is very high compared to the CV for the flow. In such cases, the CV for the computed load will be increased very little. However, when the CV for the constituent is low

decreasing the accuracy of flow measurements will significantly increase the CV for the calculated load. Reducing the accuracy of flow measuring devices (i.e. using less accurate equipment) should not be considered for a sampling point unless it can be demonstrated that the CV's for all constituents are high in relation to the CV for flow.

3.3 PRESENCE/ABSENCE DETERMINATION

The monitoring regulation for the petroleum refining sector refers to closed characterization and open characterization. The purpose of closed characterization is to identify the presence of compounds on the EMPPL that are not routinely measured. The purpose of open characterization is to identify the presence of compounds that are not on the EMPPL.

This section first explores the question of sampling frequency required to determine presence or absence of compounds in industrial effluent and then presents an example application using simulated data to demonstrate the effect of different data set characteristics on the ability of a sampling scheme to detect the presence of compounds.

3.3.1 Application of the Binomial Distribution

The question of sampling frequency requirements for determining the presence/absence of EMPPL compounds may be stated statistically as:

"What is the probability of obtaining at least one result above detection limits?"

The binomial distribution with the assumptions listed below is used as a starting point:

1. the probability of success is the same for each trial, and
2. the trials are independent.

The binomial distribution is given by the equation:

$$b(x; n, \theta) = \binom{n}{x}\theta^{x}(1-\theta)^{n-x}$$

where: x is the number of successes

n is the number of trials and

θ is the probability for success and is constant from trial to trial (Freund, 1962).

If p is the probability of getting a success then q is the probability of a failure and q = 1-p. If we chose the number of successes as zero then q will give the probability of obtaining at least one sample above detection limits. In practice, binomial probabilities are rarely calculated directly but are available from tables. Probability for various values of θ , q and N derived from tables contained in Freund (1962) are shown below.

BINOMIAL PROBABILITIES FOR DETECTING AT LEAST ONE SAMPLE ABOVE DETECTION LIMITS

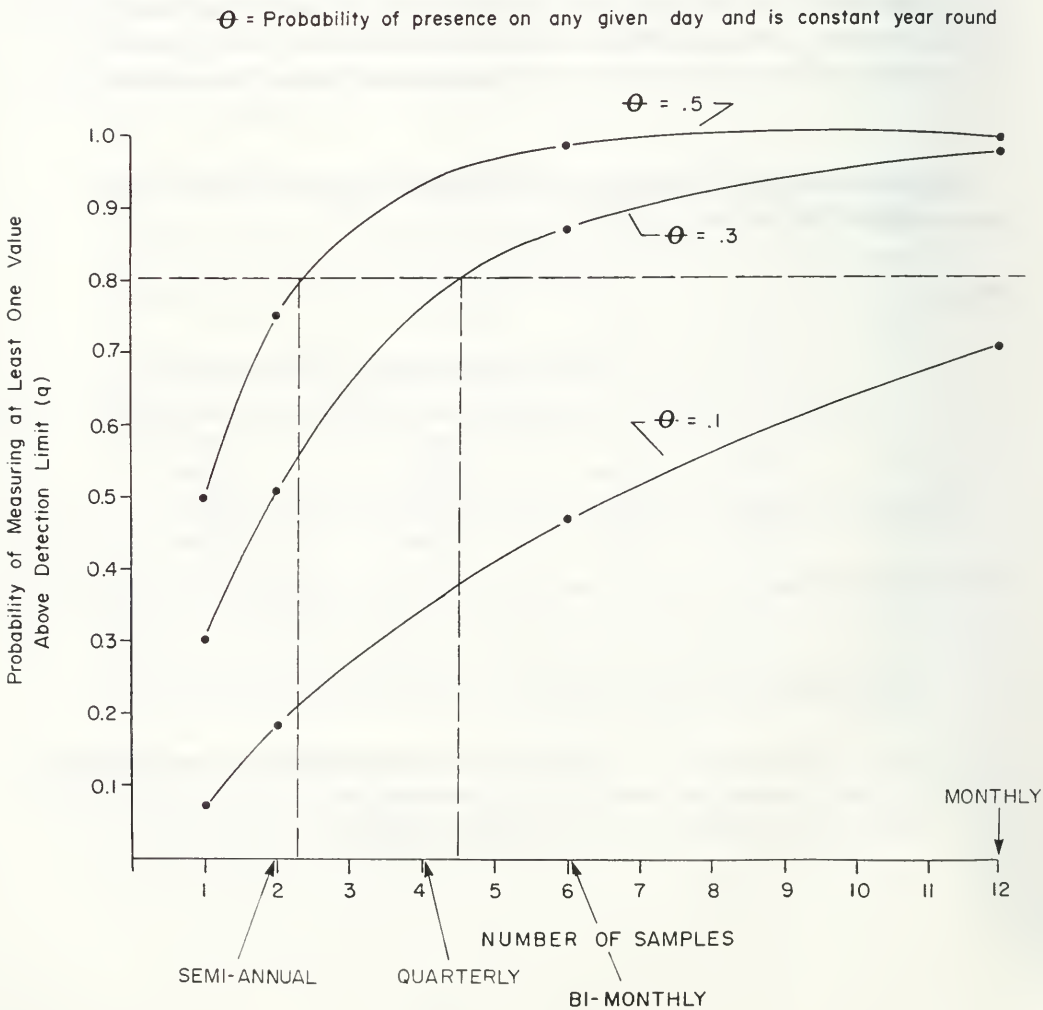
			q	
Θ	N = 12 (Monthly)	N = 6 (Bi-monthly)	N = 2 (Twice per year)	N = 1 (Once per year)
.5	.9998	.9844	.7500	.5000
.3	.9862	.8824	.5100	.3000
.1	.7176	.4686	.1900	.1000

These results are also shown in Figure 3.

If a goal of the project were to design a program that would always insure a probability of detecting at least one sample above the detection limit then Figure 3 can be used to determine the sampling frequency for different values of θ.

PROBABILITY OF AT LEAST ONE SAMPLE BEING ABOVE DETECTION LIMITS

FIGURE 3



For example, for a θ of .5, (i.e., there is a .5 probability that a compound is present on any given day and this probability is constant year round) then N be must at least 3 (e.g., sample once every four months) to yield a q of .8 (.8 probability of measuring at least one value above detection limit).

If θ is 0.3 then N must be 5 (approximately quarterly sampling) and if θ is .1 then N must be 16 (approximately once every three weeks).

The value of θ will vary from industry to industry and from parameter to parameter within an industry. The majority of the parameters on an industry's EMPPL will have values equal to 1 (always detected, e.g. suspended solids) or almost equal to 1 (e.g. chromium was above detection limits on 80% of samples thus θ = .8).

All industrial sectors (and many industries) will undergo pre-regulation monitoring for characterization purposes. Assuming that at least 4 samples are obtained from pre-monitoring the binomial distribution tells us that there is a 95% probability that at least one above detection limit value will have been obtained for θ>.5.

The parameters that are of interest in characterization are the ones with lowest θ and thus the one that are the hardest to detect (i.e. will require greater sampling frequency).

Figures 4 and 5 illustrate the effect of different values of θ for quarterly sampling and monthly sampling, respectively. Once again assuming that 0.8 is the minimum probability acceptable for the program, Figure 4 shows that quarterly sampling would be suitable for parameters which have a θ >.32. Similarly, Figure 5 shows that monthly sampling would be appropriate for parameters where θ >.12

3.3.2 Serial Correlation

The binomial distribution was used in the previous section to identify appropriate sampling frequency required to identify presence/absence of a compound. In this section, an empirical technique is used to investigate a complicating factor that is not accounted for in the application of the binomial distribution, namely; *serial correlation*.

FIG 4.BINOMIAL PROBABILITY

FOR QUARTERLY SAMPLING

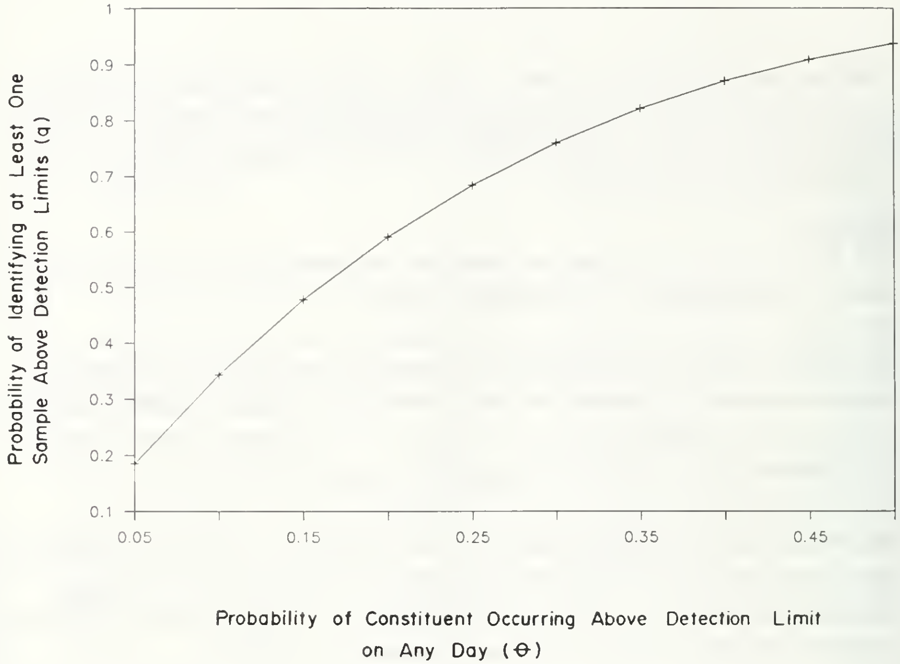
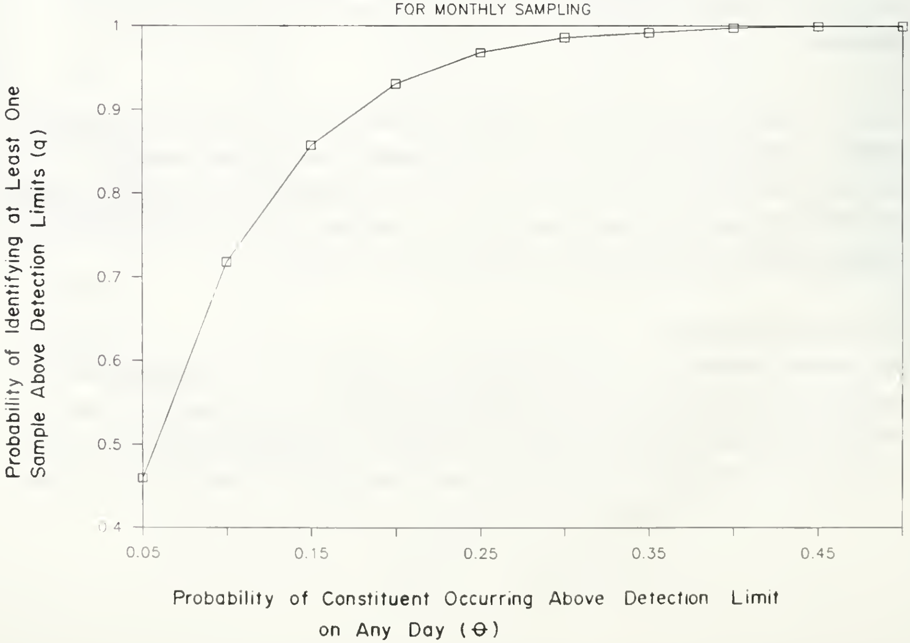


FIG 5.BINOMIAL PROBABILITY

FOR MONTHLY SAMPLING



Serial correlation (sometimes referred to as autocorrelation) is the degree to which a data set is related to itself at some specified lag period. In practical terms, this may be thought of as the correlation coefficient of a data set with itself after it has been shifted by a lag unit (often 1). Serial correlation is denoted by the autocorrelation coefficient r. Industrial effluent data often exhibit serial correlation at lag = 1.

Serial correlation was selected for further study in order to test the sensitivity of the assumption of independence in the binomial distribution.

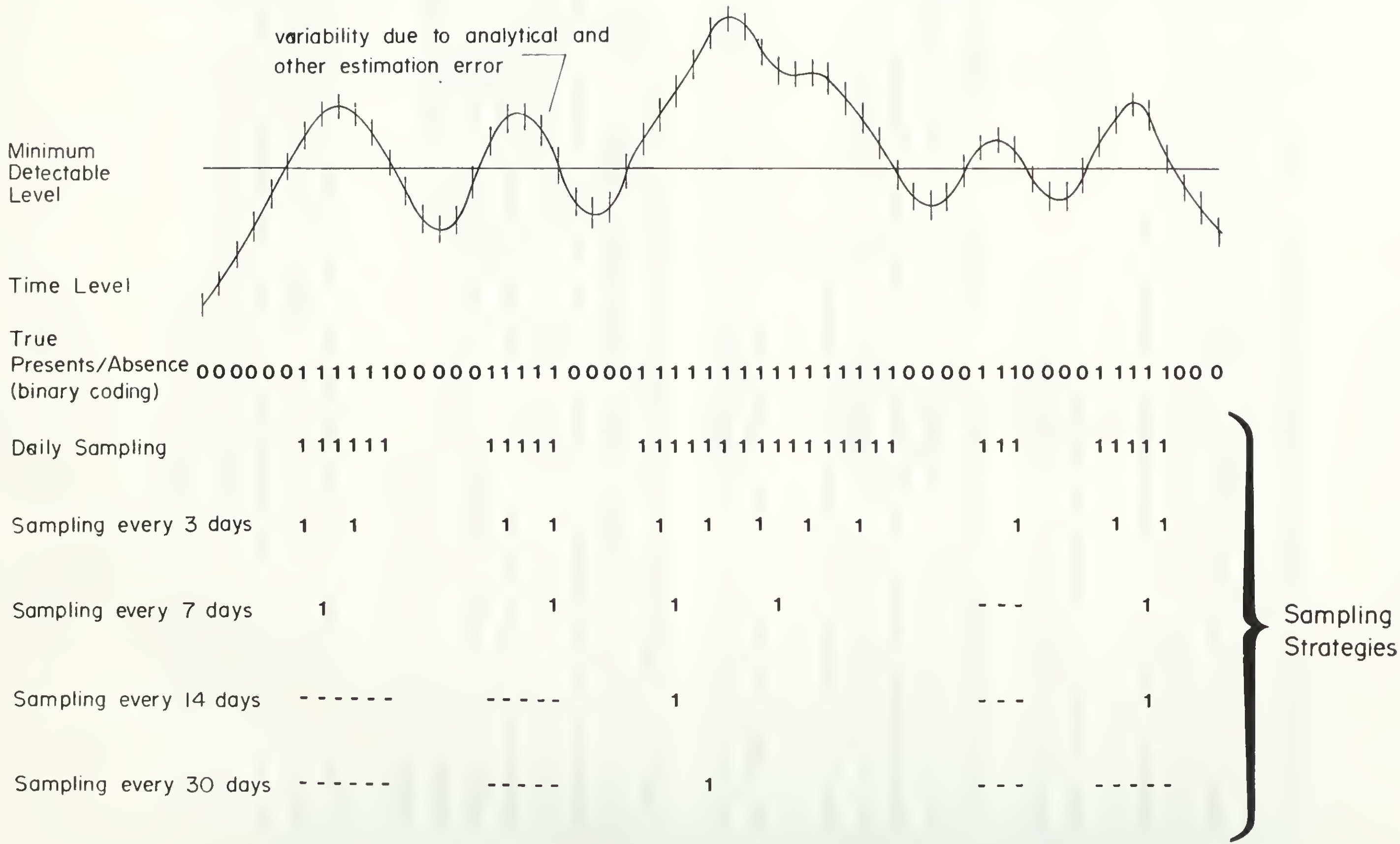
A series of MINITAB macros (Appendix B) were developed by Dr. R. Green of the University of Western Ontario to simulate and sample data sets to test the efficiency of various sampling schemes to identify above detection limit values. The results obtained were used to develop curves relating the percentage of above detection limit runs identified for various sampling schemes and autocorrelated values.

This approach was based on the use of a binary data set (i.e., a series of ones (1's) and zeros (0's) to represent the true presence (1) or the true absence (0) of a compound.

Figure 6 is a representation of a hypothetical time series in both analog and binary forms. (This is not the actual data base used in subsequent analysis, but a simplified version used here for illustration purposes.) The "true level" is represented by the thick line which can be seen to vary in magnitude with time. Regular sampling produces the series of ticks along the true level. The size of the ticks represents the combined variability associated with sampling analytical and other estimation errors. The horizontal line is the minimum detectable level. This is shown to be constant although, in reality, this too could change daily.

The binary coding relates to presence/absence rather than the true level. It may be thought of as true presence/absence. The binary data set is obtained by checking to see whether the true level is below the minimum detectable level (absent = 0) or above (present = 1). A series of 1's is called a run and may be thought of as a violation event or pollution event. A run can be length one or greater.

SIMULATED TIME SERIES
FIGURE 6



The pattern of variations in the true presence/absence can be adequately modelled by varying two parameters in a binary time series, i.e, the mean (θ) of the true presence/absence and the autocorrelation coefficient (r). (Note: θ is used in place of μ to facilitate comparison with the Binomial Distribution.) In the case of the binary data base the mean (θ) is also the probability of a violation occurring on a given day. The mean can vary from 0.0 (total absence) to 1.0 (total presence).

Autocorrelation (r) is the correlation of a time series variable with itself specifying a lag factor, in this case one. An autocorrelation coefficient of zero would indicate no relationship between one day's concentration and the next (i.e., this would satisfy the assumption of independence in the binomial distribution).

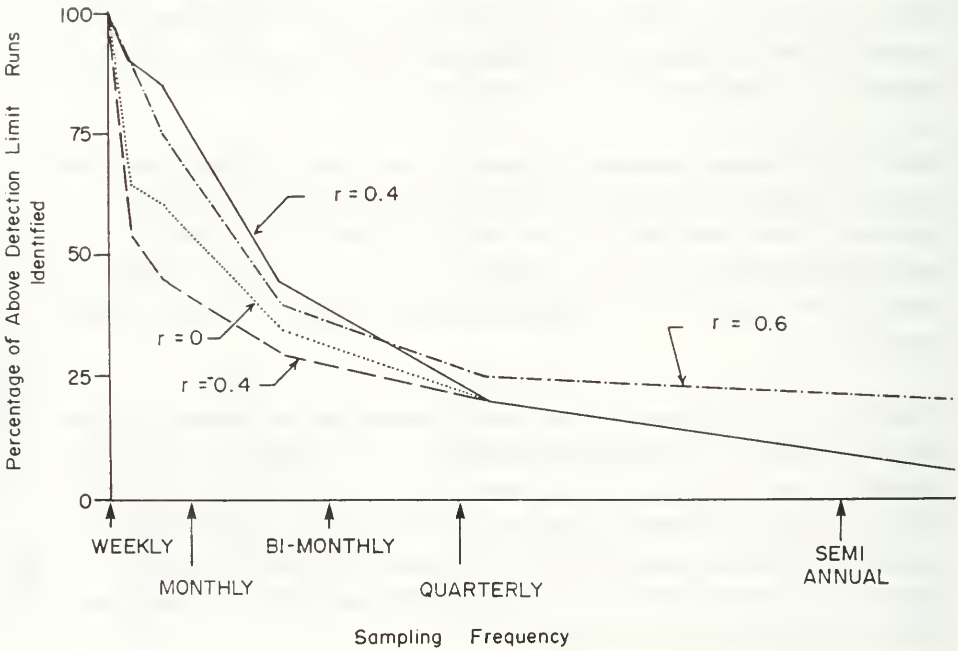
The MINITAB macros (contained in Appendix B) were used to simulate a binary data set and to examine the effects of serial correlation on the efficiency of sampling schemes to identify above detection limit values. The data were considered to represent presence/absence of a compound on a weekly basis. The total length of the simulation was 4 years (N = 210). The program uses a θ value of .5 and four different values of r.

The effects of serial correlation in the simulated data set on the efficiency of various sampling schemes in identifying above detection limit runs is shown in Figure 7. At sampling frequencies of quarterly and greater positive serial correlation resulted in increased ability to identify above detection limit runs whereas negative serial correlation resulted in decreased ability. At a frequency of twice per year an autocorrelation coefficient of +.6 increased the ability to identify runs whereas no difference was observed for other values of r.

Further simulations and use of different values of θ could be run to refine the analyses; however, it is unlikely that the general findings above would change.

THE EFFECT OF SERIAL CORRELATION ON SAMPLING EFFICENCY

FIGURE 7



3.3.3 Example Applications

Nineteen data sets were generated by the "SIMULATE" program and sampled repeatedly to test the ability of various sampling schemes to detect presence of a contaminant. The data sets sampled included both approximately *normal* (4) and *non-normal (15) distributions*. The coefficients of variations ranged from 8% to 154%. In terms of relative variability levels as defined in Section 3.2.1, the distribution was as follows:

Relative Variability	Number of Data Sets
Low	6
Medium	6
High	7
Very High	0

All data sets had positive autocorrelation at lag = 1.

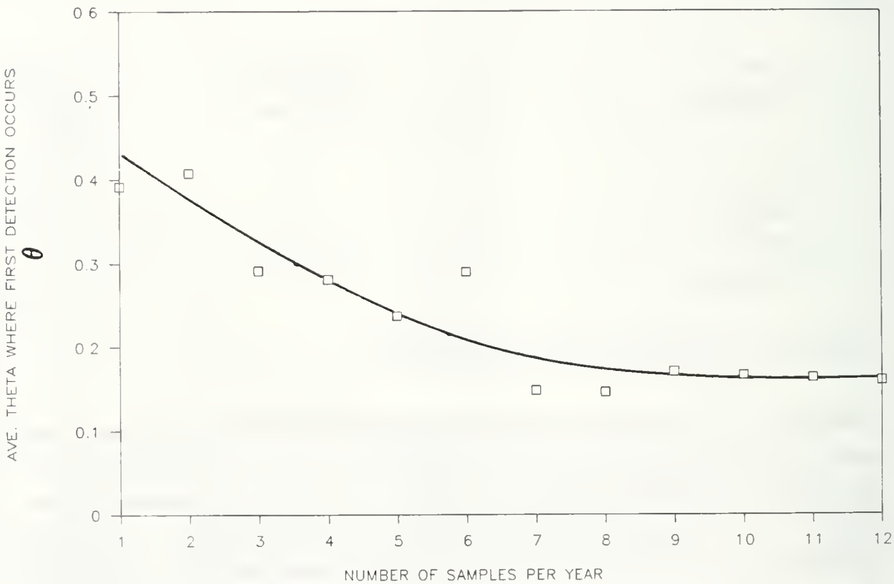
Each data set was sampled ten times for each sampling frequency and θ value and the results averaged. By changing the detection limit for a data set a new value for θ (probability of a contaminant being present)could be derived. The average number of detections per θ value and per sampling frequency was then calculated. The θ value where detection first occurred was then selected by manual methods. The time series plots and summary statistics for each data set are shown in Appendix C1 and C2, respectively.

Next a matrix was derived containing the minimum θ value that a sampling frequency was capable of detecting for each data set. Finally, the nineteen data sets were averaged and the results graphed (see Figure 8). The best fit line in the graph has been drawn "by eye".

The graph shows that the average θ value where detection first occurs declines as the number of samples obtained increases, as one would expect. Annual and semi-annual

THETA VERSUS NUMBER OF SAMPLES

FIGURE 8



sampling do not appear capable of reliably detecting presence when θ is less than 0.4. (It should be noted that these simulations did detect presence at very low θ values on occasion but not consistently.)

At sampling frequencies greater than about six times per year (or bi-monthly), the value decreases slowly to a minimum of about 0.15 for monthly (12) sampling.

Figure 8 was used to estimate the number of samples (N) required to identify at least one above detection limits for various θ values. The results are summarized below along with the associated binomial distribution probabilities (**q**) for these θ, and N values.

Θ	N	Binomial Probability (q)
.15	12	85%
.20	7	79%
.30	4	76%
.40	2	64%

3.3.4 <u>Discussion</u>

The binomial distribution appears to be a reasonable and simple method for determining sampling frequency requirements for a program to identify presence/absence for compounds. Violation of the independence assumption does not appear to significantly alter results. Serial correlation can potentially alter the ability of a sampling scheme to detect presence, for example, positive correlation enhances the ability to detect presence. Since most data sets will be positively correlated to some degree the binomial distribution would tend to be slightly conservative in this regard.

Use of the empirically derived Figure 8 to design a program and comparison of results to the binomial distribution indicated that monthly sampling would be required for θ as low as .15 and quarterly sampling would be required for θ as low as .30. Semi-annual sampling would be adequate only where θ is >.4 and a lower probability of detection is acceptable (i.e. about 64%).

4.0 GENERAL PROTOCOL FOR ESTIMATING SAMPLING FREQUENCY

The findings discussed in Chapter 3 relied on simulated data and broad classes of industrial variability levels. The next logical step to address the specific issues of sampling frequency of interest in this study would be the analysis of actual industrial effluent data using the techniques presented earlier and other methods described below.

Ward et al (1987) outlined a framework for complete design of a monitoring network that builds on previous experience. The following general protocol utilizes several of Ward's recommendations as well as techniques presented by other authors and is intended to serve as a guide for possible future investigations.

1. Review MISA program objectives and restate these in statistical terms. The methods by which BATEA effluent guidelines will be developed and how the monitoring data will be used should be specified in precise terms. When this is achieved, it should be possible to specify the required accuracy for mean monthly load estimates, etc.

2. Refine or expand relative variability levels for different industries and parameters. Identify representative industries and their associated θ , μ and CV statistics (as explained in Chapter 3).

3. Obtain suitable data bases representative of the relative variability levels and parameters identified in step #2.

4. Analyze the data base to characterize it statistically. This should include estimation of the population mean, standard deviation, seasonal variability, serial correlation and (if trend analysis or testing of statistical hypothesis is to be undertaken) identify the applicable probability distribution.

A procedure for identifying periodic trends in time series data is contained in a paper by Loftis, *et al.*, 1987. This procedure uses a correlogram to visually display periodic trends. The correlogram feature was included as an option in the DESIGN program contained in Appendix B. Several other useful graphical techniques and simple statistics that can be used to interpret environmental data are contained in a paper by Berthouex, *et al.*, 1981. These features include:

1. scattergram of the time series (included as an option in the DESIGN program),

2. histograms of absolute, relative and cumulative frequencies (included in the DESIGN program),

3. calculation of simple statistics such as mean, standard deviation, maximum, minimum, range, coefficient of variation and autocorrelation coefficient (included in the DESIGN program),

4. moving averages, and,

5. cumulative sum chart.

Many other statistical procedures are available for time series pattern definition. These features are found in most commercially available statistical analysis packages. For a description of these procedures, the reader is referred to any statistical text on time series analysis.

5. For many statistical tests (e.g. comparison of means between industrial sectors) use the information from step #4 to select the appropriate probability distribution function to match the test requirements to the population characteristics.

In hydrology (and water quality aspects of hydrology), it has been common practice to assume water constituents are normally distributed or that transformed data are normally distributed (e.g. Ward, et. al., 1979; Loftis, et. al., 1987). Under assumptions of normality subsequent analysis are simplified.

Water quality constituents may have probability density functions (PDF's) or distributions other than normal (e.g. Poisson, log-normal, negative binomial, uniform, Pearson Type III, etc.) or the PDF's may be ill defined or not defined (distribution free). In such cases, application of statistics for normal distributions will not apply.

If a known PDF can be fitted to the data then the statistics for that PDF can be used. If no fit is found then the data are said to be distribution free and non-parametric techniques are used in statistical tests.

The "SIMULATE" program includes two tests for measuring normality, i.e., *skewness* and *kurtosis*. Other measures, such as plotting the data on probability paper and Kolmogorov - Smirnov statistics are available through commercial statistics packages.

The simulated data used in the examples in this report include data with approximately normal PDF and distribution-free data.

6. Repeat the analysis described in Chapter 3.0 using actual data.

7. Based on the binomial distribution and empirically derived relationships between θ (the probability that a constituent is present above detection limits) and sampling frequency the following requirements were identified (assuming that 80%* is the minimum acceptable probability for detection).

Frequency	Suitable for θ
Monthly	> .15
Bi-monthly	> .25
Quarterly	> .35
Semi-annual	> .50 or (*at 75%) or 7.40 (*at 64%)

5.2 RECOMMENDATIONS

1. Review MISA objectives and restate these in statistical terms. Specifically, determine accuracy and precision requirements for identification of BATEA effluent limits. Also specify the minimum θ that the characterization program should detect and the minimum probability for detection that is acceptable.

2. Obtain industrial data bases and perform the analysis outlined in Chapter 3 to test the results of this study.

Respectfully submitted,

GARTNER LEE LIMITED

J.E. O'Neill, B.Sc.
Hydrologist,
Senior Consultant

6.0 REFERENCES

Berthouex, P.M. and W.G. Hunter, 1981.
Simple statistics for interpreting environmental data. Journal WPCF, Volume 53, Number 2.

Canning, R.P., 1988.
Personal communications regarding an overview assessment of industrial effluent variability. Zenon Environmental Inc., June 1988.

Freund, J.E., 1962.
Mathematical Statistics. Prentice Hall, Inc. Englewood Cliffs, New Jersey, pp. 390.

Loftis, J.C., P.S. Porter and G. Settembro, 1987.
Statistical analysis of industrial wastewater monitoring data. Journal WPCF, Vol. 59, No. 3.

Loftis, J.C., R.C. Ward and G. Smille, 1983.
Statistical models for water quality regulation. Journal WPCF, Vol. 55, No. 8.

MINITAB Inc. (Statistical Analysis Package).
3081 Enterprise Drive, State College, PA., USA.

MOE, 1988.
Effluent Monitoring Regulations for the Petroleum Refining Sector, July 1988.

Overman, R.T. and H.M. Clark, 1960.
Radioisotope Techniques, McGraw Hill, New York.

Sanders, T.G. (Editor), 1979.
Design of Water Quality Monitoring Networks. Manual for Short Course.

Sharma, A.K., D. Weatherbe, and R. Gorabedian, 1988.
MISA approach to setting limits and detecting compliance. Paper presented at the Workshop for Statistical Methods for the Assessment of Point Source Pollution, C.C.I.W., Burlington, September 13, 1988.

Spiegel, M., 1961.
Theory and Problems of Statistics. McGraw Hill Book Company, pp. 359.

Ward, R.C., J.C. Loftis and G.B. McBride, 1987.
The "Data-rich but information-poor" syndrome in water quality monitoring. Environmental Management, Vol. 10, No. 3, pp. 291-297.

Ward, R.C., J.C. Loftis, K.S. Nielsen, and R.D. Anderson, 1979.
Statistical evaluation of sampling frequencies in monitoring networks. Journal WPCF, Volume 51, No. 9.

Yevjevich, V., 1972.
Probability and Statistics in Hydrology. Water Resources Publications. Fort Collins, Colorado, pp. 302.

GLOSSARY OF TERMS

Accuracy is the closeness to the true value of the quantity being measured or to an accepted reference value.

Autocorrelation coefficient is a measure of the degree to which a data set is related to itself at some specified lag period, usually one day. Autocorrelation is sometimes termed serial correlation. The presence of autocorrelation in a data set violates the assumption of independence upon which many statistical techniques depend.

BASIC is a computer programming language.

BATEA stands for Best Available Technology that is Economically Achievable.

Coefficient of variation is the ratio of the standard deviation to the mean and is usually expressed as a percentage.

Confidence interval is the range in which the true value (e.g. the mean) is expected to fall with a certain confidence level.

Confidence interval width is the numerical extent or range of the confidence limits. For example, if the confidence limits were $\pm$ 5 mg/L then the confidence interval width would be 10 mg/L.

Confidence level is quantitative expression of the reliability of an estimated value. The expression is usually stated in probability terms.

Confidence limits are the numerical values that express the ends of the confidence interval.

Kurtosis is the degree of peakedness of a distributor, usually taken relative to a normal distribution.

Macros are a set of instructions, (e.g., in MINITAB) created by a computer user to perform specific functions (e.g., statistical calculations or mathematical manipulations. They may be thought of as a computer program written in the language of a particular software package.

Mean. The mean is a descriptor of the central tendency of a set of observations. The arithmetic mean (the most common measure) is the sum of the observations divided by the number of observations.

MINITAB is a commercially available statistical analysis package.

Non-normal refers to a probability density function that is not normal. The pdf may be some other recognized distribution or it may be undefined.

Normal or *normally distributed* refers to a probability density function (pdf) which has the characteristics of the normal pdf.

Precision is the variation in an observation or set of observations due to random error. It is the measure of the repeatability of a series of observations or measurements.

Relative error in estimating the mean is a measure of the efficiency of a sampling scheme to estimate a mean value.

Relative precision in estimating the mean is a measure of the relative width of the 95% confidence interval.

Serial correlation: See autocorrelation coefficient.

Skewness refers to the shape of a probability density function (usually compared to the normal distribution). If the longer tail occurs to the right the distribution is said to be skewed to the right or to have positive skewness.

Significance level is the maximum probability with which we would be willing to risk a Type I error when testing a given hypothesis.

Standard deviation is a descriptor of the variability of a set of observations.

Variability is the degree that a value such as effluent greatly changes over time or space.

STATISTICAL ASSESSMENT
OF SAMPLING FREQUENCY REQUIREMENTS
FOR SELECTED ASPECTS OF
THE MISA PROGRAM
VOLUME 2 OF 2

TECHNICAL APPENDICES

PREPARED FOR:

MISA ADVISORY COMMITTEE

PREPARED BY:

GARTNER LEE LIMITED

GLL 88-194

NOVEMBER, 1988

DISTRIBUTION:
CC: 1 CLIENT
2 FILE

LIST OF APPENDICES

APPENDIX A

AN OVERVIEW ASSESSMENT OF THE VARIABILITY OF INDUSTRIAL EFFLUENT QUALITY AND QUANTITY

APPENDIX A

AN OVERVIEW ASSESSMENT OF THE VARIABILITY OF INDUSTRIAL EFFLUENT QUALITY AND QUANTITY

1.0 INDUSTRIAL EFFLUENTS

Water quality variability is a key factor in the determination of sample frequency requirements. Of particular interest in this study is variability of constituents which occurs as a result of changes in effluent quality and quantity.

The range of conditions possible in effluent from Ontario industries in large and to a great extent, unknown. The characterization of industrial effluents is, in fact, one of MISA's objectives.

An overview assessment was undertaken by Zenon Environmental Inc. (Canning, 1988) to provide insight and background information concerning the range of variability likely to be encountered in industrial effluent quantity and quality.

Variability in effluent quality will range from low to high generally in response to:

- Degree of effluent treatment. Higher levels of treatment, particularly where equalization and biological treatment are included, tends to reduce effluent quality variations and/or lengthen their periodicity under normal operating conditions. Plant upsets can cause levels to increase dramatically within short periods of time (hours).

- Batch vs. continuous processing. Batch processing methods tend to increase variations in effluent quality. The smaller the batch size, the higher the frequency of these variations.

- Plant size and degree of product mix. Small plants which produce a wide range of products tend to have effluent which varies widely in quality over relatively short periods (hours or days).

Variability in effluent quantity varies from low to high generally in response to:

- Continuity of processing operation (1 shift, 2 shifts or 3 shifts per day; 5 or 7 days per week). Batch processes will generally produce high variability over shorter durations compared to continuous processes.

- Availability of equalization and/or effluent storage facilities which tend to reduce flow variations and lengthen their periodicity.

- Seasonal variations will occur at plants which have large areas of unpaved controlled surface drainage, e.g. mine sites.

- Daily and/or weekly flow variations will occur in response to rainfall events at plants which have large paved areas which contribute potentially contaminated stormwater, e.g. refineries, petrochemical plants. Extent of dampening will depend on availability of storm surge and/or equalization ponds.

Table 1 presents nine examples of industries which represent a broad range of effluent quality and quantity conditions. These broad groupings formed the basis for subsequent investigation for estimates of mean concentration and loads and presence/absence of compounds presented in Chapter 3 of Volume 1 of the report.

2.0 IDEAL DATA BASE CHARACTERISTICS

In ideal terms, a preliminary data set which can be used to determine sampling frequency requirements in a program should have the following characteristics:

- long time series,
- very frequent sampling,
- regular sampling intervals,
- good and consistent sample collection techniques
- good and consistent analysis methods,
- low and consistent detection limits, (no censored data)
- wide range of parameters, and
- on suitable electronic format

Support staff of the MISA Advisory Committee assisted in determining the status and availability of data bases that could be used as a source of "preliminary data" and form the basis for determining sampling frequency requirements. The results are shown in Table 2.

PROJECT: 88-194

Table 1: OVERVIEW ASSESSMENT OF INDUSTRIAL EFFLUENT VARIABILITY

Variability in Effluent Quality	Variability in Effluent Quantity	Example Industries	Comments
Low	Low	Petroleum Refining	Continuous 24 h/d; 7 d/wk operation plus high level of effluent treatment (inc. biological treatment) tends to produce effluents which vary over a relatively narrow range of quality and quantity compared to other industry groups.
Low	Med	Base Metal Mining	Large tailing ponds will tend to dampen variability in effluent quality with time but large areas which collect contaminated surface drainage will cause quantities to vary day to day in response to rainfall events and seasonally in response to snow melt.
Low	High	Organic Chemical Manufacturing (single product; batch processes)	Industries which produce basically the same product in batches on a repetitive basis (e.g. phenolic resin manufactures) will tend to have effluents whose quality changes very little with time but which may vary widely (hour by hour) in volume in response to batch production schedules.
Med	Low	Pulp & Paper	Continuous 24 h/d; 7 d/wk operation plus lagoon based biological treatment tends to moderate variations in effluent quality and dampen effluent flow fluctuations.
Med	Med	Inorganic Chemical Manufacturing (large, continuous processes, for high use ind. chemicals)	Larger, continuous production facilities for high use industrial chemcials, e.g. NaOH, H SO, HCl, HNO would be expected to show moderate variability in terms of both effluent quantity and quality. Organics when present tend to be at or below detectable limits. pH is main effluent control parameter. Excursions can last minutes to hours.
Med	HIgh	Inorganic Chemical (batch processes)	Generally applies to batch processes without equalization ponds or effluent holding tanks. Pattern of variability will range from low (weekly to monthly) to high (hourly or daily) depending on batch size and product mix. The larger the product mix and smaller the batch size - the more frequent the variations.
High	Low	Organic Chemical Manufacture (continuous processes)	Generally applies to mid and large size continuous processing plants which manufacture commodity and bulk chemicals, e.g. industrial solvents (aromatic, aliphatic, chlorinated).
High	Med	Organic Chemical Manufacture	Generally applies to mixed continuous and batch processes, e.g. synthetic rubbers, fibers, plastics,etc.
High	High	Organic Chemical Manufacture	Generally applies to batch produced specialty chemical manufacturing with large product mix e.g. pharmaceuticals, paints, dyes and inks, etc. Minimum effluent treatment will enhance variability of effluent quantity and quality. Quantity and quality of (indirect) effluent usually varies on a continuous basis from zero to maximum depending on activity in plant.

TABLE 2: DATA SOURCES WHICH COULD BE USED AS PRELIMINARY DATA BASES

SECTOR	MAJOR DATA SOURCES (Reports etc.)	SUMMARY	NO. OF PARAMETERS ANALYZED FOR	MDL RANGES FOR ORGANICS	SAMPLING PROFILE	STATE OF DATABASE
Petroleum Refining	Sampling and Analysis of Refinery Effluents to Assess Variations in Trace Contaminant Concentrations	pre-regulation monitoring--MISA 2 refineries influent/effluent(lagged) from biological trmt plant	30 volatiles/44 base neutral extract. 9 metals/12 conventionals 95 total	EPA Method 624 <10ppb EPA Method 625 Most <5ppb APHA Standard Methods	6 hr flow proportional composite sample 6 times every other day for 11 days (Sept. 1986)	report
	Survey of Trace Substances in Canadian Petroleum Industry Effluents of Current Information and a Monitoring Approach	analysis of plant intakes, biological influents and final plant effluents a 3 Ont. refineries	58 priority pollutants 46 samples	EPA Methods		report
	Evaluation of the Variability of Trace Organic Substances in Petroleum Refinery Effluent	analysis of 15 intake waters/15 effluents from one Sarnia refinery over a two month period	15 influent/15 effluent	organics <10ppb	24 hr composite samples over two months	report
	several other studies from PACE and API					
Organic Chemical Manufacturing	Upper Great Lakes Connecting Channel Study--1986-1988	point source survey sampling of UGLCCS parameters, conventional pollutants and the USEPA Priority Pollutants	USEPA Priority Pollutants List total of 105 parameters	volatiles (1-5 ug/L) extractables (2-24 ug/L)	1-6 day (5 day ave.) sampling during fall of 1986	computer spreadsheet report
	MISA Pilot Studies	effluent monitoring phase--sampling on 9 pipes -Polysar 72", cole drain, biox and Dow 1st, 2nd, 3rd, & 4th St. sewer	40 organics	?	2 times/week for ten months beginning May 15th	computer spreadsheet report
	Pre-regulation monitoring a Industry analysis	3 to 6 days consecutive sampling of selected pipes including the influent and major process streams	USEPA 129/11 conventionals +dioxins and furans	various depending on lab -<10ppb	sequential sampling from 3 to 6 days	spreadsheet
	b MOE scans	1 scan on major process streams	74 metals-scan 68 organics		one sample	hardcopy
Other Sectors	Pre-regulation monitoring see above various miscellaneous sources					Pulp & Paper Sector data entered onto spreadsheet

APPENDIX B

DOCUMENTATION OF COMPUTER PROGRAMS FOR INVESTIGATION OF SAMPLING FREQUENCY REQUIREMENTS

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1.0 INTRODUCTION

This appendix contains documentation for the set of computer programs developed for the report: "Statistical Assessment of Sampling Frequency Requirements for Selected Aspects of the MISA Program". For convenience, the complete set of programs and data files are referred to by the name "DESIGN" (a reference to the design of water quality monitoring programs).

DESIGN consists of two parts. The first part is a stand-alone executable program written in Microsoft Quick Basic Version 4.0. The second part is a series of MINITAB macro files. MINITAB is a commercially available statistical analysis package. Macro files are actually data sets which contain instructions for the MINITAB programs. The two parts of DESIGN are complementary. For example, data sets generated by the BASIC program can be analyzed by MINITAB Macros. The programs allow for generation and analysis of water quality concentration data with prescribed statistical characteristics, e.g., well-defined (approximately normal) distributions versus ill-defined (non-normal) distributions, low versus high variability and different detection limits. The programs also allow for input of data from external sources for analysis, e.g., real data generated through the MISA program or elsewhere.

Program options for the first part are accessed through a menu system as shown in Figure 1. Both program parts give the user instructions interactively.

FIGURE 1: MENUS IN THE "SIMULATE" PROGRAM

a) Main Menu

```
MENU
________________________________________

1) CHANGE DEFAULT SETTINGS
2) DRAW POLLUTOGRAPH
3) LOAD DATASET
4) CALCULATE FREQUENCY DISTRIBUTION
5) CALCULATE SUMMARY STATISTICS
6) WRITE DATA TO FILE
7) REDRAW POLLUTOGRAPH
8) ANALYSIS PROGRAMS
9) QUIT
________________________________________

 1  2  3  4  5  6  7  8  9  ->
```

b) Options Menu

```
MENU
________________________________________

1) INITIALIZE PRINTER SETTINGS
2) SET TINY PRINT
3) DRAW DELAY FACTOR
4) RETURN TO MAIN MENU
________________________________________

 1  2  3  4  ->
```

c) Analysis Menu

```
MENU
________________________________________

1) CONFIDENCE INTERVAL FOR MEANS
2) PRESENCE/ABSENCE OF CONSTITUENTS
3) CORRELOGRAM
4) CONFIDENCE INTERVAL (FULL)
5) RETURN TO MAIN MENU
________________________________________

 1  2  3  4  5  ->
```

2.0 PURPOSE

The main objective of this set of computer programs is to generate monitoring scenarios which demonstrate the following:

- the effects of standard deviations on estimates of mean concentrations and confidence intervals,

- the effects of detection limits on identifying presence or absence of constituents.

Since only limited data are available at present for analyses, a secondary objective of the model was to generate artificial data sets of predetermined characteristics which can be used for analysis and to gain insight on sampling frequency requirements.

3.0 HARDWARE AND SOFTWARE REQUIREMENTS

3.1 BASIC PROGRAMS

The following minimum equipment and software are required to run the BASIC programs.

An IBM PC or equivalent computer with:

- one DS/DD 360 K Byte drive
- one Hard Disk (must be designated "C" Drive)
- colour/graphics display adapter
- colour or composite monitor
- 256 K Byte memory
- DOS Version 3.0 or higher
- graphics printer

The following are not required but will enhance program performance and/or function.

- IBM AT (or equivalent)
- math coprocessor
- joystick and games serial port
- LOTUS 1-2-3 Version 2.0 or greater

3.2 MINITAB MACROS

The minimum requirement to run this set of programs is the MINITAB statistical analysis program. The hardware and other requirements to run MINITAB are specified in their product literature.

4.0 INSTALLATION

One DS/DD diskette labelled "DESIGN PROGRAMS (88-194)" is included.

The diskette contains three sub-directories as follows:

DESIGN - (Part 1) contains BASIC programs and supporting files.

MTAB - (Part 2) contains MINITAB macros.

DATA - simulated Data sets used in the report.

To load Part 1, the BASIC programs, follow these instructions:

1. Turn on computer in the normal way.

2. Determine if there is a sub-directory under the root directory of drive C named "DESIGN". If there is, rename it.

3. Once you are sure there is no sub-directory called "DESIGN"
   type A:
   type CD/DESIGN
   type INSTALL

4. The installation program will create a directory C:\DESIGN and copy the relevant files into it.

5. The current drive will be C:\DESIGN. To run the program type SIMULATE. Specific user instructions are provided in Section 5.0 of this Appendix.

Part 2, MINITAB macros can be copied into your MINITAB subdirectory as follows:
(Note: you must have the MINITAB program to continue)

1. Type CD C:\MTAB (assuming MTAB is the subdirectory containing MINITAB)

2. Type Copy A:\MTAB*.*

5.0 USER INSTRUCTIONS

5.1 BASIC PROGRAMS

5.1.1 Data Preparation

The DESIGN - Part 1 program can generate artificial data or import external files for analysis in various ways. Before proceeding with any session you must either already have data for analysis, or you must generate it.

5.1.1.1 Import Data Files

To import data it must be in an ASCII file in the C:\DESIGN sub-directory and named:

DATA00??.PRN

"??" may be any two letter A - Z or number 0 - 9 combination. No distinction is made on the case of the letter, thus a maximum of 1296 data sets are permitted at any one time. (If an existing data set name is given when the program asks for it the old data set will be overwritten).

The information in the data set must represent daily conditions (e.g., mean, min, or max). The data set consists of 365 rows (or records) containing numerical values (in decimal format) which represent daily concentrations. No delimiters, e.g., "," are allowed. Example data are included on the distribution disk in the file: DATA00NPRN. To obtain a printout of the data enter TYPE DATA00N.PRN>PRN. This data set is used in the verification process discussed later.

If you wish to input a data file which does not meet the above requirements it will be necessary for the user to modify the original data into suitable form. For example, if there are missing data, these should be estimated. If there are more than 365 values then a subset of 365 values should be selected. If multiple values exist for any day, these should be averaged and the average values entered.

If the data file to be input meets the above requirements it can be loaded for analysis by typing in the "RUN #" (i.e., 0-9 or A-Z) at the program request.

If the data file does not exist you will be returned to the previous menu. If the data file exists but has incorrect format an error will be produced and the program will terminate.

EXAMPLE SESSION

		(comments)
Type	C:\DESIGN	change directories
	SIMULATE	this starts the program
	3	select input data file from main menu
	1	select run #1, i.e., DATA001.PRN.
	7	display of x-y graph of file
	8	selects further analysis programs
	etc.	follow program prompts.

5.1.1.2 Generate Data Files

An important feature of the program is that the user is provided with the option of generating artificial data for analysis. This feature requires a joystick and games adapter to be installed in the computer.

This option is accessed through the main menu (see Figure 1) (enter 2). The program asks for the maximum concentration, units and run number. Next the program draws an x-y graph, sounds an alarm and waits for the user to press a key to start input. The y-

axis is the concentration of the constituent which is to be simulated. The x axis is time, in 12 numbered increments representing months.

When a key is pressed the program moves the cursor steadily from left to right. The magnitude of the concentration is determined by the Y-axis positioning of the joystick control. The rate of change of concentration is determined by the horizontal speed of the cursor and the speed with which the user changes the joystick position.

With a little practice a wide range of time-concentration graphs can be easily generated.

<u>NOTE</u>: When using an AT computer and/or when a math co-processor is installed the horizontal cursor speed may be too fast. The speed can be modified by selecting 1 from the main menu (change options) and 3 from the options menu (change cursor speed). To slow the cursor speed increase the delay factor.

It is possible to enter and/or edit data sets manually using any available word processor.

When 365 data values have been generated the program pauses for the user to view the graph.

If desired, a printout of the graph on the screen can be obtained. To accomplish this the user must have entered the DOS "graphics" command prior to running the program. If you have not done this and wish to print the screen, stop the program now. At the DOS prompt type: C:\DOS\GRAPHICS (assuming your DOS programs are in a \DOS directory).

To printout the graph type Shift-Prt Scr.

When generating artificial data the user may not get what is wanted on the first try. The program allows the user to redraw the graph as many times as desired simply by entering the "3" - (Draw) option. If the graph looks promising further analyses are provided for.

Option 4 on the main menu invokes a frequency distribution analysis of the data on the screen. This provides a visual impression of the type of distribution generated. Shift Prt Scr will print the frequency distribution if a graphics printer is installed. Output from this option is shown in Figure 2.

Option 5 (Summary statistics) calculates the mean, maximum, minimum, standard deviation, variance, range, skewness coefficient, Kurtosis coefficient, excess coefficient and autocorrelation coefficient (lag = 1) for the data on the screen. Output from this option is shown in Figure 3.

Option 7 redraws the graph on the screen.

If the user is satisfied with the graph produced it can now be saved by selecting option 6. The program prompts for a run # (0-9 and/or A-Z) which will be used to name and save the data on disk.

5.1.2 Data Analysis

At this point it is assumed that a suitable data set exists on file. Analysis programs are available at two levels.

Option 4 (frequency distribution), 5 (summary status) and 7 (redraw graph) which were described in the previous section are still available for use. If you have not reviewed the data for some time or are inputing new data (option 3) it may be desirable to do these analyses first.

The main analysis programs are accessed through option 8 on the main menu (Figure 1). This loads the analysis option menu which produces the following choices.

1. Confidence Limits
2. Presence/Absence Analysis
3. Correlogram
4. Confidence Limits (Full)
5. Return to Main Menu

FIGURE 2: OUTPUT FROM OPTION 1-4

a)

CLASS	
0 - 14	4
15 - 29	11
30 - 44	40
45 - 59	49
60 - 74	73
75 - 89	66
90 - 104	56
105 - 119	42
120 - 134	16
135 - 154	8

A B S O L U T E F R E Q U E N C Y D I S T R I B U T I O N

b)

CLASS		CUMULATIVE FREQUENCY
0 - 14	1.10	1.10
15 - 29	3.01	4.11
30 - 44	10.96	15.07
45 - 59	13.42	28.49
60 - 74	20.00	48.49
75 - 89	18.08	66.58
90 - 104	15.34	81.92
105 - 119	11.51	93.42
120 - 134	4.38	97.81
135 - 154	2.19	100.00

R E L A T I V E F R E Q U E N C Y D I S T R I B U T I O N

c)

CLASS	CUMULATIVE FREQUENCY
0 - 14	1.10
15 - 29	4.11
30 - 44	15.07
45 - 59	28.49
60 - 74	48.49
75 - 89	66.58
90 - 104	81.92
105 - 119	93.42
120 - 134	97.81
135 - 154	100.00

C U M U L A T I V E F R E Q U E N C Y D I S T R I B U T I O N

FIGURE 3 SCREEN OUTPUT FOR OPTION 1-5

```
D E S C R I P T I V E   S T A T I S T I C S

THE MEAN IS                               76.315
THE SD IS                                 28.691
THE MINIMUM IS                             0.000
THE MAXIMUM IS                           154.000
RANGE IS                                 154.000
THE VARIANCE OF X IS                     823.197
THE SKEWNESS COEFFICIENT IS              0.0069
THE KURTOSIS COEFFICIENT IS              2.5471
THE EXCESS COEFFICIENT IS               -0.4529
THE COEFFICIENT OF VARIATION IS          37.6
AUTOCORR. COEFFICENT (LAG=1) IS          0.0755
```

5.1.2.1 Confidence Limits Analysis

Upon selecting this option the program asks for the run # (data set name). You must provide a valid entry - even if you have been analyzing a data set previously.

The program divides the 365 (daily) data set into 12 subsets (months) of data consisting of 30 days each. (The last 5 days of the year are ignored). For each month the following is computed: upper confidence limit, mean, lower confidence limit, range, standard deviation, variance, autocorrelation coefficient (lag = 1), skewness coefficient, delta and frequency.

This is done for three subsets of the month as follows:

1. daily (N=30) assumed to be the population (true) data
2. thrice weekly (N = 13), and
3. weekly (N = 4)

The results of analysis are reported in a printout. (See Figure 4)

When executing the analysis the user is given the option of printing out the raw data.

For each analysis run 12 simulations (i.e., months) are performed.

5.1.2.2 Presence/Absence Analysis

This option (#2) performs a function similar to the MINITAB macros. It differs in the way that the binary data set is derived. Whereas the MINITAB macros generate a binary data set according to a specified θ and r this program module uses data values (i.e., concentrations) and determines presence or absence of a constituent based on a detection limit supplied by the user.

(θ is the true proportion of detected values to total number of days, i.e., population mean and r is the autocorrelation co-efficient with a lag = 1).

FIGURE 4: OUTPUT OF OPTION 8-1

```
* * * * CONFIDENCE INTERVAL ANALYSIS * * * * *
      [illegible]
SIMULATION RUN NUMBER [illegible]
[illegible]-1988
[illegible]
STATISTICS CALCULATED USING POPULATION MEAN [illegible]

DATASET [illegible]:\DESIGN\DATA00N.PRN
```

MONTH

EP	N		MEAN	UCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	21.08	20.[illegible]	801.[illegible]	[illegible].15	-0.[illegible]	[illegible]	0.0[illegible]	1
[illegible]	[illegible]	[illegible]	[illegible]	41.[illegible]	32.22	[illegible]	[illegible]	[illegible]	-0.[illegible]	[illegible]	-1.[illegible]	2
[illegible]	[illegible]	[illegible]	[illegible].25	55.[illegible]	32.99	10.34	[illegible]	-0.[illegible]	[illegible]	[illegible]	10.61	[illegible]

MONTH

EP	N		MEAN	UCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	34.[illegible]	1[illegible]	[illegible]	[illegible]	[illegible]	[illegible].00	1
[illegible]	[illegible]	[illegible]	[illegible]	64.[illegible]	34.1[illegible]	26.25	[illegible]	[illegible]	[illegible]	-1.4[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	55.[illegible]	30[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

MONTH

EP	N		MEAN	UCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	22.[illegible]	4[illegible]	[illegible]	[illegible]	[illegible]	-10.6[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

MONTH

EP	N		MEAN	UCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
[illegible]	[illegible]	[illegible]	[illegible]	5[illegible]	1[illegible]	2[illegible]	[illegible]	[illegible]	[illegible]	-1.2[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-0.[illegible]	-10.[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

MONTH

EP	N		MEAN	UCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-0.5[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-1.[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-1.54	[illegible]	[illegible]

MONTH

EP	N		MEAN	UCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
[illegible]	[illegible]	[illegible]	[illegible]	51.[illegible]	26.[illegible]	22.0[illegible]	[illegible]	0.2[illegible]	[illegible]	-1.7[illegible]	[illegible]	[illegible]

FIGURE 4 page 2

The program samples the population data set for a variety of sampling scenarios as follows:

- annual
- semi-annual
- quarterly
- bi-monthly
- monthly
- weekly
- thrice weekly
- daily

The following statistics are calculated for each sampling scenario: mean, standard deviation, variance, number of hits, number of misses, % of actual, FOD and FOD error.

The number of hits is the number of 1's or above detection limit values identified by sampling. The % of actual is the number of hits as a percentage of the true number of 1's. FOD is the frequency of detection and represents the estimated percent of samples that are above detection based on sampling. FOD error is the difference between the estimated FOD and population FOD. The population statistics are provided in the "daily" sampling scenario.

The program offers the opportunity to repeat the analyses as often as desired using different user-supplied detection limits. Changing the detection limit will change θ. Sample output using the test data is shown in Figure 5.

5.1.2.3 Correlogram

Option 3 performs a correlogram analysis of the data set for lag = 1 to 180. A correlogram is a useful tool for investigation of periodic behaviour in time series data (Loftis, et. al., 1987). It is a plot of the estimated autocorrelation coefficient for the time series data. Regular peaks in the autocorrelation function usually indicate

FIGURE 5: OUTPUT FROM OPTION 8-2

```
* * * * [illegible] PROGRAM * * * *
        [illegible]
        RUN DATE [illegible]
        [illegible]

[illegible]

[illegible]

[illegible]
```

periodic behaviour. If periodicity is identified (e.g., weekly cycles) then the data should be adjusted to remove the cycles.

This program calculates autocorrelation values and writes the results to a file named:

C:\DESIGN\AUTOCOR?.PRN
(Where "?" is user-supplied)
This file can be imported into LOTUS 1-2-3 using the File-Import command sequence. LOTUS 1-2-3 can be used to analyze the data further or produce a plot of the correlogram. See Figure 6 for an example of a correlogram.

5.1.2.4 Confidence Limit (Full Analysis)
Option 4 performs the same analysis as Option 1 except the sampling frequency varies from 1 to 14. See Figure 7 for sample output.

Option 5 ends the analysis module and returns the user to the main menu.

Data files used or generated by these programs can easily be input to other package programs for further analysis. For example, the data file C:\DESIGN\DATA001, PRN could be input to LOTUS 1-2-3 for further analysis. Similarly, the data set could be input to any statistical analysis program for analyses using more sophisticated statistical analysis.

To accomplish this the user should refer to the documentation for the particular analysis program to be used.

5.2 MINITAB MACROS
As discussed previously the user must own a copy of the statistical analysis package "MINITAB" to run this part of the model. The MINITAB macros are contained on the distribution diskette in the files.

FIGURE 6 EXAMPLE OUTPUT FROM OPTION 8-3*

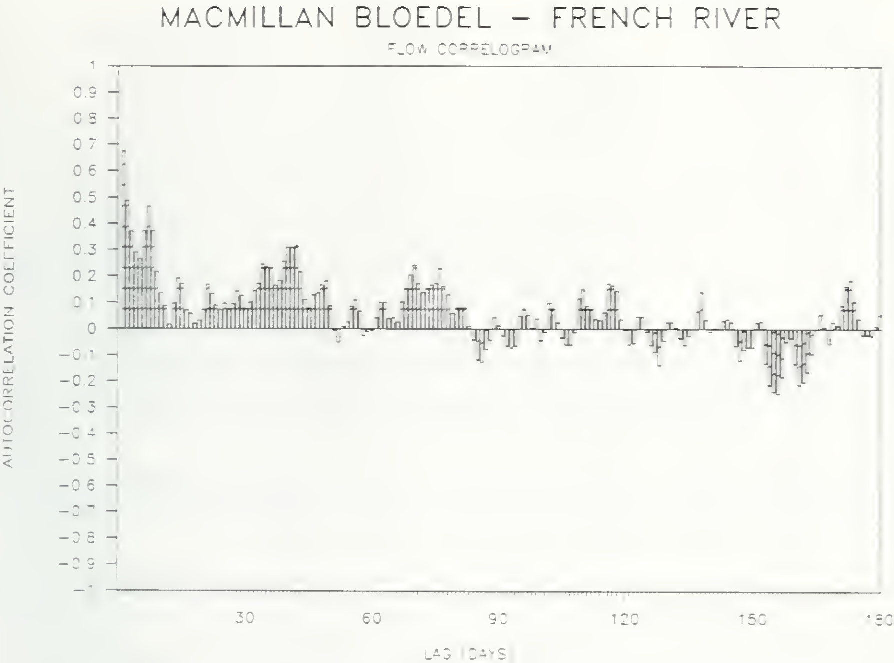


FIGURE 7: EXAMPLE OUTPUT FROM OPTION 8-4

```
* * * *CONFIDENCE INTERVAL ANALYSI. * * * *
     F U L L   A N A L [illegible] S I S
     GARTNER LEE LIMITED
 IMULATION RUN NUMBER [illegible]
09-1[illegible]-198[illegible]
[illegible]
```

DATASET [illegible] DESTIN DATAUON.PEN

STATISTICS CAL[illegible] ATED USING POPULATION MEAN (M) FOR MONTH 1

[illegible]	[illegible]	MAX	UCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
[illegible]	[illegible]	[illegible]	54.65	21.05	28.32	801.85	0.1[illegible]	-0.4[illegible]	-0.7[illegible]	[illegible]	1
	[illegible]	[illegible]	45.19	[illegible]	23.09	57[illegible].27	[illegible]	-1.30	1.24	-11.15	2
	[illegible]	[illegible]	[illegible]	41.57	2[illegible].0[illegible]	444.20	0.24	[illegible]	[illegible]	0.61	3
		[illegible]	45.17	72.5[illegible]	17.40	[illegible]	-1.45	1.40	[illegible]	[illegible]	4
	[illegible]	[illegible]	[illegible]	66.5[illegible]	32.67	1067.64	0.02	-1.25	[illegible]	-6.44	5
	[illegible]	[illegible]	[illegible]	34.91	34.15	1166.1[illegible]	-0.6[illegible]	-0.02	[illegible]	[illegible]	6
[illegible]		[illegible]	55.[illegible]	[illegible]	10.74	[illegible]	[illegible]	1.77	[illegible]	10.61	7
	[illegible]	[illegible]	15.[illegible]	68.44	17.79	[illegible]	-0.21	[illegible]	-1.46	-19.22	8
	[illegible]		[illegible]	[illegible]	[illegible]	[illegible]	.42	[illegible]	[illegible]	[illegible]	[illegible]
			[illegible]	[illegible]	[illegible]	304.[illegible]	-1.5[illegible]	[illegible]	[illegible]	-2.2[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	451.78	24.02	[illegible]	-0.54	-1.7[illegible]	[illegible]	-30.98	11
	[illegible]			[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	20.4[illegible]	12
			[illegible]	[illegible]	17.5[illegible]	307.54	[illegible]	1.41	-1.06	[illegible]	13
			[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-0.7[illegible]	14
		[illegible]	[illegible]	40[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	-1.[illegible]	[illegible]	15
		[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

[illegible] MEAN [illegible] FOR MONTH [illegible]

[illegible]	[illegible]	MAX	UCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]		[illegible]	45.75	24.37	22.06	486.46	-0.11	-0.37	1.05	-11.15	[illegible]
[illegible]	[illegible]		43.[illegible]	[illegible]	[illegible]	[illegible]		[illegible]	[illegible]	[illegible]	[illegible]
	[illegible]	[illegible]	[illegible]	[illegible]	17.66	[illegible]	[illegible]	1.50	[illegible]	[illegible]	[illegible]
	[illegible]	[illegible]	[illegible]	[illegible]	32.40	1050.[illegible]	[illegible]	-0.8[illegible]	-0.[illegible]	-6.44	[illegible]
			[illegible]	[illegible]	[illegible]	1150.1[illegible]	[illegible]	-0.4[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	60.[illegible]	24.[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
	[illegible]		[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
	[illegible]		[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]		[illegible]	17.[illegible]	[illegible]	1[illegible].1[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
		[illegible]			[illegible]		-1.[illegible]				[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	26.50	702.[illegible]	-1.0[illegible]	[illegible]	[illegible]	20.40	[illegible]
		[illegible]	[illegible]	224.[illegible]	12.50	156.[illegible]	-1.0[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
		[illegible]		[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
		[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	1.[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
	[illegible]		[illegible]		[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

[illegible]

BINSER0.DAT
BINSER1.DAT
BINSER2.DAT
BINSER3.DAT
BINSER4.DAT
BINSER5.DAT
BINSER6.DAT

These macros consist of two parts; namely, a part to generate a binary data set with specified θ and r (BINSER0.DAT to BINSER2.DAT) and a part to analyze the data set for various sampling scenarios (BINSER3.DAT to BINSER6.DAT).

5.2.1 Data Generation

To generate a binary (0/1) series (i.e., time series) having specified parameters type:

EXEC "BINSER0.DAT"

The program prompts the user to input the mean (θ) of the series, i.e., the true long-term mean probability of state 1 occurring, the autocorrelation coefficient (r) with lag = 1 and the length (N) of the series to be generated.

To accomplish this, user must type:

let K1 = 0.5 (return)
let K2 = 0.6 (return)
let K3 = 90 (return)
EXEC "BINSER1" (return)

In the above example, θ = 0.5, r = 0.6 and N = 90.

The program produces the specified binary data set in column C10.

5.2.2 Data Analysis

To analyze a binary data set generated previously, type:

"EXEC "BINSER3.DAT"

The program will then print the binary data set, plot presence/absence versus N and calculate mean and number of runs of 1.

Next, the 1's are replaced by run number and printed. A run is a number of 1's occurring sequentially without interruption.

Next, the autocorrelation coefficient (lag = 1) and chi-square statistic (1 degree of freedom) are calculated and printed.

At this point the program prompts the user to enter the sampling interval to investigate. To sample every other day, for example, type:

LET K14 = 2
EXEC "BINSER5.DAT"

The program now samples at the specified frequency and calculates and prints the number of runs detected. The sampling frequency can be repeated for various intervals as necessary.

5.2.3 Other Features

The macros BINSER3.DAT to BINSER6.DAT can be used either to analyze a data set which has been generated by BINSER0.DAT to BINSER2.DAT or data from some other source, e.g., simulated data from the BASIC programs (Section 5.1) or real data. In the case of real data it will be necessary to translate the actual concentration values into a series of 1's and 0's based on some user-specified criteria such as detection limit. This must be done manually or with some other program not currently available. The results must be stored in column C10 for the MINITAB macros to work.

The analysis can then be performed (using BINSER3.DAT) as before to calculate θ and r and then analyze for various sampling scenarios.

Alternatively, the user could write his own macros which could be used to analyze the data and determine θ and r. Those values could then be used to generate a binary data set, with these characteristics, for further analysis.

When running these macros the user must select an appropriate time frame, e.g., daily or weekly. If N=210 is specified and a time frame of daily observations is assumed then the programs would be useful for examining sampling schemes in the range from every other day to monthly, (i.e., you cannot get two semi-annual samples in a data base of less than a year). If a time frame of weekly is assumed for a value of N=210 then sampling schemes in the range of weekly to semi-annually can be evaluated.

When selecting a time frame the user is cautioned that he must also choose θ and r values that are consistent with the selected time frame. For example, for daily samples θ is the true mean daily probability of presence and r is the autocorrelation coefficient for lag = 1 day and for weekly samples the appropriate weekly statistics must be used.

6.0 PROGRAM VERIFICATION

6.1 INTRODUCTION

The purpose of program verification is to document proof that the program works as intended. Two methods have been used to perform verification checks. The first method is to perform all analysis by hand and compare the results to the program output. The second method is the use of other programs (known to function correctly) to provide a check of results.

The data set used for verification of the BASIC programs was C:\DESIGN\DATA00N\PRN and is contained on the distribution disk.

6.2 FREQUENCY DISTRIBUTION

The frequency distribution program module is accessed through Option #4 on the main menu of the SIMULATE program. Following is a listing of the verification steps used (i.e., Method 1).

MENU ITEM #4 (FREQUENCY DISTRIBUTION)

1. The data to be analyzed must be input by means of the Main Menu Item #3 (run number = N)

2. Select number of intervals = 10

3. Determine Maximum from Table 1: = 154 for Record #49

4. Determine Minimum from Table 1: = 0 for Record #56

5. Range = Maximum-Minimum = 154-0 = 154

6. Interval Width = Range/Interval number = 154/10 = 15.4 (program truncates result therefore Interval width = 15).

Gartner Lee Limited

CALCULATIONS

SHEET ___ OF ___

CALCULATED BY	CHECKED BY	PROJECT NO	88-194
DATE. [illegible]	DATE.	PROJECT NAME	M.A.C
		DESCRIPTION	"SIMULATE.EXE"

TABLE 1 PRINT OUT OF DATA USED FOR PROGRAM VERIFICATION

c:\DESIGN\DATA00N.PRN

1	36 -3	2	32 -3	3	28 -2	4	67 -5	5	0 -1
6	4 -1	7	79 -6	8	46 -4	9	68 -5	10	39 -3
11	32 -3	12	105 -8	13	90 -7	14	68 -5	15	92 -7
16	60 -5	17	108 -8	18	62 -5	19	99 -7	20	66 -5
21	80 -6	22	58 -4	23	105 -8	24	52 -4	25	86 -6
26	65 -5	27	82 -6	28	62 -5	29	102 -7	30	83 -6
31	44 -3	32	99 -7	33	55 -4	34	117 -8	35	80
36	74 -5	37	104 -7	38	45 -4	39	114 -8	40	78
41	72 -5	42	98 -7	43	60 -5	44	120 -9	45	81
46	96 -7	47	113 -8	48	78 -6	49	154 ← 10 MAX	50	143
51	78 -6	52	84 -6	53	40 -3	54	18 -2	55	56
56	0 ← -1 MIN	57	57 -4	58	46 -4	59	41 -3	60	44
61	94 -7	62	27 -2	63	82 -6	64	60 -5	65	72
66	48 -4	67	72 -5	68	84 -6	69	50 -4	70	118
71	112 -8	72	66 -5	73	108 -8	74	70 -5	75	114
76	80 -6	77	97 -7	78	88 -6	79	82 -6	80	87
81	74 -5	82	102 -7	83	44 -3	84	109 -8	85	54
86	101 -7	87	66 -5	88	97 -7	89	76 -6	90	78
91	98 -7	92	29 -2	93	106 -8	94	94 -7	95	70
96	110 -8	97	54 -4	98	105 -8	99	66 -5	100	74
101	98 -7	102	42 -3	103	101 -7	104	43 -3	105	93
106	52 -4	107	66 -5	108	31 -3	109	66 -5	110	64
111	28 -2	112	50 -4	113	20 -2	114	73 -5	115	35
116	66 -5	117	94 -7	118	36 -3	119	100 -7	120	60
121	83 -6	122	88 -6	123	102 -7	124	109 -8	125	60
126	82 -6	127	63 -5	128	90 -7	129	66 -5	130	77
131	96 -7	132	56 -4	133	102 -7	134	56 -4	135	101
136	37 -3	137	87 -6	138	42 -3	139	92 -7	140	32
141	82 -6	142	63 -5	143	78 -6	144	78 -6	145	41
146	95 -7	147	42 -3	148	101 -7	149	42 -3	150	87
151	40 -3	152	87 -6	153	47 -4	154	92 -7	155	62
156	88 -6	157	54 -4	158	60 -5	159	67 -5	160	37
161	80 -6	162	34 -3	163	106 -8	164	45 -4	165	96
166	40 -3	167	98 -7	168	47 -4	169	101 -7	170	34

Gartner Lee Limited

CALCULATIONS

SHEET ___ OF ___

CALCULATED BY: J. Kroneel	CHECKED BY:	PROJECT NO. 88-144
DATE: Sept 9/88	DATE:	PROJECT NAME: M.A.C.
		DESCRIPTION: "SIMULATE. EYE"

TABLE 1 - continued

171	89-6	172	48-4	173	82-6	174	28-2	175	66-5
176	40-3	177	60-5	178	54-4	179	34-3	180	93-7
181	34-3	182	102-7	183	40-3	184	104-7	185	43-3
186	78-6	187	50-4	188	62-5	189	80-6	190	47-4
191	76-6	192	54-4	193	124-9	194	107-8	195	84-6
196	125-9	197	95-7	198	104-7	199	54-4	200	35-3
201	60-5	202	122-9	203	54-4	204	106-8	205	129-9
206	130-9	207	135-10	208	122-9	209	69-5	210	70-5
211	33-3	212	1-1	213	50-4	214	41-3	215	49-4
216	104-7	217	104-7	218	80-6	219	98-7	220	117-8
221	64-5	222	108-8	223	136-10	224	140-10	225	114-8
226	70-5	227	85-6	228	116-8	229	82-6	230	98-7
231	72-5	232	81-6	233	118-8	234	62-5	235	88-6
236	119-8	237	88-6	238	63-5	239	109-8	240	83-6
241	66-5	242	115-8	243	82-6	244	40-3	245	99-7
246	108-8	247	121-9	248	137-10	249	137-10	250	68-5
251	104-7	252	94-7	253	62-5	254	74-5	255	107-8
256	63-5	257	45-4	258	84-6	259	57-4	260	66-5
261	96-7	262	50-4	263	76-6	264	65-5	265	50-4
266	74-5	267	88-6	268	68-5	269	56-4	270	78-6
271	120-9	272	84-6	273	88-6	274	109-8	275	54-4
276	50-4	277	95-7	278	106-8	279	59-4	280	79-6
281	52-4	282	60-5	283	79-5	284	80-6	285	55-4
286	75-6	287	94-7	288	64-5	289	82-6	290	69-5
291	68-5	292	78-6	293	43-3	294	85-6	295	67-5
296	86-6	297	111-8	298	65-5	299	91-7	300	58-4
301	107-8	302	66-5	303	98-7	304	60-5	305	112-8
306	67-5	307	98-7	308	68-5	309	80-6	310	88-6
311	64-5	312	114-8	313	52-4	314	108-8	315	115-8
316	46-4	317	85-6	318	123-9	319	103-7	320	41-3
321	114-8	322	92-7	323	40-3	324	118-8	325	92-7
326	52-4	327	126-9	328	72-5	329	71-5	330	124-9
331	48-4	332	118-8	333	59-4	334	83-6	335	50-4
336	96-7	337	108-8	338	27-2	339	71-5	340	84-6
341	31-3	342	19-2	343	27-2	344	95-7	345	78-6
346	46-4	347	116-8	348	128-9	349	129-9	350	80-6
351	48-4	352	91-7	353	112-8	354	57-4	355	40-3
356	69-5	357	121-9	358	137-10	359	132-9	360	74-5
361	68 -	362	40 -	363	22 -	364	67-5	365	64-5

Note: 362 | 40 —3

record # | value | frequency class

TABLE 1 - continued

DATASET c:\DESIGN\DATA00n.PRN

```
    36    32    28    67     0     4    79    46    68    39    32   105    90
  68    92    60   108    62    99    66    80    58   105    52    86    65
82    62   102    83    44    99    55   117    80    74   104    45   114    78
    72    98    60   120    81    96   113    78   154   143    78    84    40
  18    56     0    57    46    41    44    94    27    82    60    72    48
72    84    50   118   112    66   108    70   114    80    97    88    82    87
    74   102    44   109    54   101    66    97    76    78    98    29   106
  94    70   110    54   105    66    74    98    42   101    43    93    52
66    31    66    64    28    50    20    73    35    66    94    36   100    60
    83    88   102   109    60    82    63    90    66    77    96    56   102
  56   101    37    87    42    92    32    82    63    78    78    41    95
42   101    42    87    40    87    47    92    62    88    54    60    67    37
    80    34   106    45    96    40    98    47   101    34    89    48    82
  28    66    40    60    54    34    93    34   102    40   104    43    78
50    62    80    47    76    54   124   107    84   125    95   104    54    35
    60   122    54   106   129   130   135   122    69    70    33     1    50
  41    49   104   104    80    98   117    64   108   136   140   114    70
85   116    82    98    72    81   118    62    88   119    88    63   109    83
    66   115    82    40    99   108   121   137   137    68   104    94    62
  74   107    63    45    84    57    66    96    50    76    65    50    74
88    68    56    78   120    84    88   109    54    50    95   106    59    79
    52    60    79    80    55    75    94    64    82    69    68    78    43
  85    67    86   111    65    91    58   107    66    98    60   112    67
98    68    80    88    64   114    52   108   115    46    85   123   103    41
   114    92    40   118    92    52   126    72    71   124    48   118    59
  83    50    96   108    27    71    84    31    19    27    95    78    46     1
16   128   129    80    48    91   112    57    40    69   121   137   132    74
    68    40    22    67    64
```

7. Interval classes are therefore:

Class #			
1	0	-	14
2	15	-	29
3	30	-	44
4	45	-	59
5	60	-	74
6	75	-	89
7	90	-	104
8	105	-	119
9	120	-	134
10	135	-	154

(See Figure 4a to verify screen output is correct).

Note: Last class is larger than 15 to correct for decimal truncation in step #6.

. Count absolute frequency distribution occurrence in selected classes (limits are included in classes)

Class				Number of Values
1	0	-	14	4
2	15	-	29	11
3	30	-	44	40
4	45	-	59	49
5	60	-	74	73
6	75	-	89	66
7	90	-	104	56
8	105	-	119	42
9	120	-	134	16
10	135	-	154	8
				365

(Counts from Table 1 agree with screen output in Figure 4a)

Note: (See Table 1)

9. Compute Relative Frequency by dividing absolute frequency (each class) by 365 (total number of values) x 100 (percent)

Class				Absolute		Relative	Cumulative
1	0	-	14	4	/ <365 x 100 =	1.10	1.10
2	15	-	29	11		3.10	4.11
3	30	-	44	40		10.96	15.07
4	45	-	59	49		13.42	28.49
5	60	-	74	73		20.00	48.49
6	75	-	89	66		18.08	66.58
7	90	-	104	56		15.34	81.92
8	105	-	119	42		11.51	93.42
9	120	-	134	16		4.38	97.81
10	135	-	154	8		2.19	199.99
						100 (rounded)	

The above data agree with the screen results shown in Figure 2b and 2c.

6.3 DESCRIPTIVE STATISTICS

The descriptive statistics program is accessed through Option #5 on the main menu of the SIMULATE program. The screen output is shown in Figure 3.

The output for this program was verified by using the commercially available statistical analysis program STATPAC GOLD version 3.0. The same data set was input to STATPAC and analyzed. The results are contained in Figures 8 and 9. All of the statistics produced on Figure 3 agree with those on Figures 8 and 9. The Excess Coefficient is not calculated by STATPAC. It is simply (Kurtosis - 3) or (2.5471 - 3) = - 0.4529.

6.4 CONFIDENCE INTERVAL FOR MEANS

The confidence interval for means program is accessed through Option #8 of the main menu and Option #1 of the analysis menu (SIMULATE program). Figure 4 contains the output of the analysis.

This section of code is used repeatedly in various parts of the program to compute statistics using data which are sampled according to different scenarios.

FIGURE 8 USE OF STATPAC GOLD VERSION 3.0 TO VERIFY "SIMULATE.EXE" PROGRAM

```
USE OF STATPAC GOLD VERSION 3.0 TO VERIFY "SIMULATE.EXE" PROGRAM

DESCRIPTIVE STATISTICS FOR [illegible]DESIGN\DATADON.FRN

[illegible]

     Minimum [illegible]
     Maximum [illegible]
     Range [illegible]
     Sum [illegible]
     [illegible]
     [illegible]
     [illegible]
     [illegible]
     Standard deviation [illegible]
     [illegible] mean [illegible]
     [illegible] confidence interval [illegible] from mean [illegible]
     [illegible]
     Standard deviation [illegible]
     [illegible] [illegible]
     [illegible] [illegible]
     [illegible]

[illegible]
[illegible]
[illegible]
```

FIGURE 9 USE OF STATPAC GOLD VERSION 3.0 TO VERIFY "SIMULATE.EXE" PROGRAM

Simple Correlation Analysis

FIGURE 10: STATPAC OUTPUT FOR FIRST 30 RECORDS OF TEST DATA

```
StatPac Gold Statistical Analysis Package

DESCRIPTIVE STATISTICS FOR DATA00_.DAT

concentration
-------------

     Minimum                    =  0

     Maximum                    =  108

     Range                      =  108

     Sum                        =  1956

     Mean                       =  65.2000

     Median                     =  66.5000

     Mode                       =  Multi-Modal

     Variance                   =  801.8933

     Standard deviation         =  28.3177

     Standard error of the mean  =  5.2585

     95 Percent confidence interval around the mean =  54.8934 - 75.5066

     Variance (unbiased)                =  829.5448

     Standard deviation (unbiased)      =  28.8018

     Skewness         = -0.4861

     Kurtosis         =  2.6267

     Kolmogorov-Smirnov statistic for normality =  0.5605

Valid cases       = 30
Missing cases     = 0
Response percent = 100.0 %
```

The verification consists of two parts:

1. Checking the accuracy of statistical calculations, and
2. checking the accuracy of the sampling procedure.

The calculations were checked for month #1 of the printout and for the "daily" sampling scheme. Table 1 (Section 6.2) contains the raw data.

The sampling scenarios were derived manually and input for analysis by STATPAC. The results are shown in Figures 7 and 8. All of the statistics produced agree with those on Figure 6.

6.5 MINITAB MACROS

Following is a listing of the program output from sample data included on the diskette.

```
MTB > exec 'binser3.dat'      ***** Entry from keyboard *****

MTB > #
MTB > # You may be executing this macro as a follow-up to 'BINSER1.DAT'
MTB > #   which would have been used to generate a binary series with
MTB > #   specified parameters and to store it in column C10. If so, you wi
ll
MTB > #   interpret the following analysis and evaluation as a description
of a
MTB > #   simulated series which has known (specified) parameters. Thus the
MTB > #   sample estimates of the parameters (from 'BINSER3.DAT') can be
MTB > #   compared to their known values (the input to 'BINSER1.DAT').
MTB > #
MTB > # Alternatively, the binary series (which must be stored in column C1
0)
MTB > #   could be a real monitoring data set where state 1 represents
MTB > #   "detectable" or "violation of the standard" and state 0 represent
s
MTB > #   "below detectable" or "not in violation of the standard". If thi
 is
MTB > #   the case, then the following analysis and evaluation is for data
MTB > #   whose parameters are not known but for which estimates are desir
d.
MTB > #   You may wish to use these parameter estimates (of mu and rho )
MTB > #   as input to 'BINSER1.DAT', to simulate binary series having these
MTB > #   properties, and then to run 'BINSER3.DAT' again to evaluate the
MTB > #   efficiency of various sampling schemes for use in sampling such
MTB > #   monitoring data.
MTB > #
MTB > # Now we will analyze the binary series:
MTB > PRINT C10

C10
   1    0    0    0    1    1    1    1    1    1    0    0    1    0    0
   1    0    1    0    1    0    0    1    0    0    0    0    1    0    0
   1    0    0    0    0    0    1    1    0    0    0    0    0    0    0
   1    1    0    0    1    1    0    1    1    1    0    1    0    1    0
   0    0    0    1    0    0    1    1    1    0    1    0    0    0    1
   0    1    0    0    1    0    0    1    0    1    0    1    1    0    1
   0    0    0    1    1    1    1    1    1    1    1    0    0    0    1
   0    0    0    0    1    1    0    1    1    0    1    1    1    0    0
   1    1    0    0    0    0    0    1    1    1    0    1    1    0    0
   1    1    0    0    0    0    1    0    0    0    0    0    1    0    1
   0    0    0    0    0    0    0    0    0    0    0    1    1    1    0
   1    0    0    1    1    1    0    1    0    0    0    1    0    1    1
   1    0    1    0    1    1    0    1    1    1    1    1    0    1    1
   1    1    0    0    1    0    0    0    1    0    0    0    1    [illegible]    1

MTB > TSPLOT C10

      -
   1.050+
         - 1   567890  3  6 8 0  3    8  1     78         67  01 345 7 9
```

```
C10       -
          -
          -
   0.700+
          -
          -
          -
          -
   0.350+
          -
          -
          -
          -
   0.000+  234       12 45 7 9 12 4567 90 23456  9012345  89  2   6 8 0
          +---------+---------+---------+---------+---------+---------+
          0        10        20        30        40        50        60


   1.050+
          -    4  789 1   5 7  0  3 5 78 0   45678901   5    01 34 678
C10       -
          -
          -
   0.700+
          -
          -
          -
          -
   0.350+
          -
          -
          -
          -
   0.000+ 123 56   0 234 6 89 12 4 6  9 123        234 6789  2  5   90
          +---------+---------+---------+---------+---------+---------+
         60        70        80        90       100       110       120


   1.050+
          - 12     890 23  67    2     8 0            234 6  901 3   7 90
C10       -
          -
          -
   0.700+
          -
          -
          -
          -
   0.350+
          -
          -
          -
          -
```

```
    0.000+   34567   1   45   8901 34567 9 12345678901   5 78   2 456 8
          +----------+----------+----------+----------+----------+----------+
         120        130        140        150        160        170        180


    1.050+
         - 1 3 56 89012 4567  0   4   8 0
C10      -
         -
         -
    0.700+
         -
         -
         -
         -
    0.350+
         -
         -
         -
         -
    0.000+  2 4  7     3    89 123 567 9
          +----------+----------+----------+
         180        190        200        210

MTB > MEAN C10
   MEAN    =      0.44762
MTB > RUNS 0.5 C10

    C10

    K =      0.5000


    THE OBSERVED NO. OF RUNS = 101
    THE EXPECTED NO. OF RUNS = 104.8476
    94 OBSERVATIONS ABOVE K  116 BELOW
              THE TEST IS SIGNIFICANT AT  0.5905
              CANNOT REJECT AT ALPHA = 0.05

MTB > LET C13(1)=C10(1)
MTB > LET K12=1
MTB > LET K13=2
MTB > LET K3=COUNT(C10)
MTB > LET K8=K3-1
MTB > OH=0
MTB > EXEC 'BINSER4.DAT' K8 TIMES
MTB > LET C13(K13)=C10(K13)-C10(K12)
MTB > LET K12=K12+1
MTB > LET K13=K13+1
MTB > LET C13(K13)=C10(K13)-C10(K12)
MTB > LET K12=K12+1
MTB > LET K13=K13+1
MTB > LET C13(K13)=C10(K13)-C10(K12)
MTB > LET K12=K12+1
```

```
MTB > LET K13=K13+1
          :                          \
          :                           \  Repeated executions
          :                           /     deleted here
          :                          /
MTB > LET C13(K13)=C10(K13)-C10(K12)
MTB > LET K12=K12+1
MTB > LET K13=K13+1
MTB > LET C13(K13)=C10(K13)-C10(K12)
MTB > LET K12=K12+1
MTB > LET K13=K13+1
MTB > OH=24
MTB > CODE (-1)0 C13,C13
MTB > PARSUM C13,C14
MTB > LET C15=C10*C14
MTB > LET K12=MAXI(C15)
MTB > #
MTB > # The number of runs of 1 is:
MTB > PRINT K12
K12      51.0000
MTB > #
MTB > # Here is the binary series again, with the 1's replaced by the run
MTB > #   number (1st, 2nd, ---  run of 1's):
MTB > PRINT C15

C15
    1     0     0     0     2     2     2     2     2     2     0     0
  3
    0     0     4     0     5     0     6     0     0     7     0     0
  0
    0     8     0     0     9     0     0     0     0     0    10    10
  0
    0     0     0     0     0     0    11    11     0     0    12    12
  0
   13    13    13     0    14     0    15     0     0     0     0    16
  0
    0    17    17    17     0    18     0     0     0    19     0    20
  0
    0    21     0     0    22     0    23     0    24    24     0    25
  0
    0     0    26    26    26    26    26    26    26    26     0     0
  0
   27     0     0     0     0    28    28     0    29    29     0    30
 30
   30     0     0    31    31     0     0     0     0     0    32    32
 32
    0    33    33     0     0    34    34     0     0     0     0    35
  0
    0     0     0     0    36     0    37     0     0     0     0     0
  0
    0     0     0     0     0    38    38    38     0    39     0     0
 40
   40    40     0    41     0     0     0    42     0    43    43    43
  0
   44 [illegible] 45    45     0    46    46    46    46    46     0    47
```

```
47
   47    47     0     0    48     0     0     0    49     0     0     0
50
    0    51

MTB > ACF 5 C10

ACF of C10

           -1.0 -0.8 -0.6 -0.4 -0.2  0.0  0.2  0.4  0.6  0.8  1.0
             +----+----+----+----+----+----+----+----+----+----+
  1   0.031                           XX
  2   0.047                           XX
  3   0.051                           XX
  4   0.086                           XXX
  5  -0.025                          XX

MTB > COPY C10 C12;
SUBC>  OMIT 1:1.
MTB > COPY C10 C11;
SUBC>  OMIT K3:K3.
MTB > CORR C11 C12,M1

Correlation of C11 and C12 = 0.031

MTB > COPY M1 C12-C13
MTB > LET K9=C13(1)
MTB > LET K10=K8*K9*K9
MTB > # The autocorrelation coefficient and the chi-square (1 df) are:
MTB > PRINT K9 K10
K9        0.0313311
K10       0.205163
MTB > #
MTB > # Now we only observe at intervals (e.g. the series is daily and yo
MTB > #   are sampling every so many days).
MTB > #
MTB > # LET K14 = The sampling (observation) interval you want, e.g. 3 if
ou
MTB > #   want every 3 days, and then enter "EXEC 'BINSER5.DAT' ".

MTB > let k14=2                    ***** Entry from keyboard *****

MTB > exec 'binser5.dat'           ***** Entry from keyboard *****

MTB > #
MTB > # The sampling interval is now:
MTB > PRINT K14
K14       2.00000
MTB > LET K9=ROUND(K3/K14+0.5)
MTB > LET K15=K14-1
MTB > SET C12
MTB > END
MTB > STACK C12 1,C12
MTB > OH=0
MTB > EXEC 'BINSER6.DAT' K9 TIMES
```

```
MTB > STACK C12 C11,C11
MTB > STACK C12 C11,C11
MTB > STACK C12 C11,C11
         |            \
         |             \     Deletion of repeated commands here
         |             /
         |            /
MTB > STACK C12 C11,C11
MTB > STACK C12 C11,C11
MTB > OH=24
MTB > COPY C11 C11;
SUBC>  USE 1:K3.
MTB > LET C12=C15*C11
MTB > NAME C15 'SERIES',C11 'OBSERVE',C12 'DETECTED'
MTB > #
MTB > # Following are the generated series (col.1), observation times (col.
2),
MTB > #   and occurrences of 1 that are detected (col.3):
MTB > PRINT C15 C11 C12

 ROW  SERIES  OBSERVE  DETECTED

   1       1        0         0
   2       0        1         0
   3       0        0         0
   4       0        1         0
   5       2        0         0
   6       2        1         2
   7       2        0         0
   8       2        1         2
   9       2        0         0
  10       2        1         2
  11       0        0         0
  12       0        1         0
  13       3        0         0
  14       0        1         0
  15       0        0         0
  16       4        1         4
  17       0        0         0
  18       5        1         5
  19       0        0         0
  20       6        1         6
  21       0        0         0
  22       0        1         0
  23       7        0         0
  24       0        1         0
  25       0        0         0
  26       0        1         0
  27       0        0         0
  28       8        1         8
  29       0        0         0
  30       0        1         0
  31       9        0         0
  32       0        1         0
  33       0        0         0
```

34	0	1	0
35	0	0	0
36	0	1	0
37	10	0	0
38	10	1	10
39	0	0	0
40	0	1	0
41	0	0	0
42	0	1	0
43	0	0	0
44	0	1	0
45	0	0	0
46	11	1	11
47	11	0	0
48	0	1	0
49	0	0	0
50	12	1	12
51	12	0	0
52	0	1	0
53	13	0	0
54	13	1	13
55	13	0	0
56	0	1	0
57	14	0	0
58	0	1	0
59	15	0	0
60	0	1	0
61	0	0	0
62	0	1	0
63	0	0	0
64	16	1	16
65	0	0	0
66	0	1	0
67	17	0	0
68	17	1	17
69	17	0	0
70	0	1	0
71	18	0	0
72	0	1	0
73	0	0	0
74	0	1	0
75	19	0	0
76	0	1	0
77	20	0	0
78	0	1	0
79	0	0	0
80	21	1	21
81	0	0	0
82	0	1	0
83	22	0	0
84	0	1	0
85	23	0	0
86	0	1	0
87	24	0	0
88	24	1	24

89	0	0	0
90	25	1	25
91	0	0	0
92	0	1	0
93	0	0	0
94	26	1	26
95	26	0	0
96	26	1	26
97	26	0	0
98	26	1	26
99	26	0	0
100	26	1	26
101	26	0	0
102	0	1	0
103	0	0	0
104	0	1	0
105	27	0	0
106	0	1	0
107	0	0	0
108	0	1	0
109	0	0	0
110	28	1	28
111	28	0	0
112	0	1	0
113	29	0	0
114	29	1	29
115	0	0	0
116	30	1	30
117	30	0	0
118	30	1	30
119	0	0	0
120	0	1	0
121	31	0	0
122	31	1	31
123	0	0	0
124	0	1	0
125	0	0	0
126	0	1	0
127	0	0	0
128	32	1	32
129	32	0	0
130	32	1	32
131	0	0	0
132	33	1	33
133	33	0	0
134	0	1	0
135	0	0	0
136	34	1	34
137	34	0	0
138	0	1	0
139	0	0	0
140	0	1	0
141	0	0	0
142	35	1	35
143	0	0	0

144	0	1	0
145	0	0	0
146	0	1	0
147	0	0	0
148	36	1	36
149	0	0	0
150	37	1	37
151	0	0	0
152	0	1	0
153	0	0	0
154	0	1	0
155	0	0	0
156	0	1	0
157	0	0	0
158	0	1	0
159	0	0	0
160	0	1	0
161	0	0	0
162	38	1	38
163	38	0	0
164	38	1	38
165	0	0	0
166	39	1	39
167	0	0	0
168	0	1	0
169	40	0	0
170	40	1	40
171	40	0	0
172	0	1	0
173	41	0	0
174	0	1	0
175	0	0	0
176	0	1	0
177	42	0	0
178	0	1	0
179	43	0	0
180	43	1	43
181	43	0	0
182	0	1	0
183	44	0	0
184	0	1	0
185	45	0	0
186	45	1	45
187	0	0	0
188	46	1	46
189	46	0	0
190	46	1	46
191	46	0	0
192	46	1	46
193	0	0	0
194	47	1	47
195	47	0	0
196	47	1	47
197	47	0	0
198	0	1	0

```
 199        0         0          0
 200       48         1         48
 201        0         0          0
 202        0         1          0
 203        0         0          0
 204       49         1         49
 205        0         0          0
 206        0         1          0
 207        0         0          0
 208       50         1         50
 209        0         0          0
 210       51         1         51

MTB > LET C16=C11*C15
MTB > #
MTB > # The number of runs of 1 that were missed by a given sampling scheme
 can
MTB > #   be calculated as the number of runs of 1 minus the number of non-
zero
MTB > #   categories listed in the following TALLYs of the data.
MTB > TALLY C16

     C16  COUNT          C16  COUNT
       0    162           34      1
       2      3           35      1
       4      1           36      1
       5      1           37      1
       6      1           38      2
       8      1           39      1
      10      1           40      1
      11      1           43      1
      12      1           45      1
      1[illegible]     [illegible]           46      3
      16      1           47      2
      17      1           48      1
      21      1           49      1
      24      1           50      1
      25      1           51      1
      26      4           N=    210
      28      1
      29      1
      30      2
      31      1
      32      2
      33      1

MTB > #
MTB > # You can change the value stored in K14 and "EXEC 'BINSER5.DAT  " ag
ain.
```

7.0 PROGRAM LISTINGS

```
DECLARE SUB AN3 ()
DECLARE SUB Script ()
DECLARE SUB READIN3 (arraydat!())
DECLARE SUB PROC2 (arraydat!(), samplenum%)
DECLARE SUB CORRELOGRAM ()
DECLARE SUB An2 ()
DECLARE SUB readin2 (arraydat!())
DECLARE SUB An1 ()
DECLARE SUB READIN (arraydat!())
DECLARE SUB Thrice (arraydat!(), matrix!(), samplenum%)
DECLARE SUB Proc1 (matrix!(), samplenum%)
DECLARE SUB Sample (arraydat!(), matrix!(), samplenum%)
DECLARE SUB Options2 (anals$())
DECLARE SUB Draw2 ()
DECLARE SUB Fileload (arraydat!())
DECLARE SUB Screensave ()
DECLARE SUB Refresh ()
DECLARE SUB Options (OPTS$())
DECLARE SUB Printfile (arraydat())
DECLARE SUB waiter ()
DECLARE SUB Ender ()
DECLARE SUB Draw1 ()
DECLARE SUB Bargraph ()
DECLARE SUB FRAME (LEFTCOL%, RIGHTCOL%, TOPROW%, BOTTOMROW%)
DECLARE SUB Menu (MENUCHOICES$(), NUMcHOsen%)
DECLARE SUB Summarize ()


'***************************************************************************
'**                  S I M U L A T E . E X E                              **
'**                       J.E. O'Neill                                    **
'**                    SEPTEMBER 3, 1988.                                 **
'***************************************************************************


OPTION BASE 1
COMMON SHARED x, Nx, n, Direct, arraydat(), xbar, xsd, Xvar, R, Flag1
COMMON SHARED Flag3, Flagjk, DELAY, MIN, max, simlength, Flagfile, Max.conc
COMMON SHARED LPS, MONTH, xc, rn$, MU, matrix(), FREQ(), excess, V, FLAG6
COMMON SHARED Hit, Missed, Samplecount, Num1, Val1, Pres1, skew, Kurt
ON ERROR GOTO HANDLER
SCREEN 0                                  ' ***FOR CGA 640  BY 200
DIM arraydat(366)                         ' ***HOLDS ONE YEAR'S DATA
DIM MEN$(9)                               ' ***MAIN MENU ITEMS
DIM FREQ(10)
DIM matrix(31)
DIM SHARED t975(30)

t975(1) = 12.71
t975(2) = 4.3
t975(3) = 3.18
t975(4) = 2.78
t975(5) = 2.57
t975(6) = 2.45
t975(7) = 2.36
t975(8) = 2.31
t975(9) = 2.26
t975(10) = 2.23
t975(11) = 2.2
```

```
t975(12) = 2.18
t975(13) = 2.16
t975(14) = 2.14
t975(15) = 2.13
t975(16) = 2.12
t975(17) = 2.11
t975(18) = 2.1
t975(19) = 2.09
t975(20) = 2.09
t975(21) = 2.08
t975(22) = 2.07
t975(23) = 2.07
t975(24) = 2.06
t975(25) = 2.06
t975(26) = 2.06
t975(27) = 2.05
t975(28) = 2.05
t975(29) = 2.04
t975(30) = 2.04
true% = -1
false% = 0
FOR I% = 1 TO 9                         ' ***READ IN MAIN MENU ITEMS
        READ MEN$(I%)
NEXT I%
DATA    1 CHANGE DEFAULT SETTINGS
DATA    2 DRAW POLLUTOGRAPH
DATA    3 LOAD DATASET
DATA    4 CALCULATE FREQUENCY DISTRIBUTION
DATA    5 CALCULATE SUMMARY STATISTICS
DATA    6 WRITE DATA TO FILE
DATA    7 REDRAW POLLUTOGRAPH
DATA    8 ANALYSIS PROGRAMS
DATA    9 QUIT
' ==========O P T I O N S ==========================================
        DIM OPTS$(4)                    ' ***OPTIONS MENU
        FOR I% = 1 TO 4
                READ OPTS$(I%)
        NEXT I%
DATA 1 INITIALIZE PRINTER SETTINGS
DATA 2 SET TINY PRINT
DATA 3 DRAW DELAY FACTOR
DATA 4 RETURN TO MAIN MENU
'  =========================================================
        DIM anals$(5)
        FOR I% = 1 TO 5
                READ anals$(I%)
        NEXT I%
DATA 1 CONFIDENCE INTERVAL FOR MEANS
DATA 2 PRESENCE/ABSENCE OF CONSTITUENTS
DATA 3 CORRELOGRAM
DATA 4 CONFIDENCE INTERVAL (FULL)
DATA 5 RETURN TO MAIN MENU
'==========================================================
OK% = false%
Flag1 = 0: Flag3 = 0
LOOP1:
        CALL Menu(MEN$(), CHOICE%)
        IF CHOICE% = 1 THEN CALL Options(OPTS$())
        IF (CHOICE% = 2) THEN
                Flag1 = 1
```

```
        CALL Draw1
END IF
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
IF (CHOICE% = 3) THEN
        CALL Fileload(arraydat())
        Flag1 = 1
END IF
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
IF ((CHOICE% = 4) AND (Flag1 = 0)) THEN
        CLS
        LOCATE 10, 20
        PRINT "YOU MUST DRAW A GRAPH BEFORE YOU CAN ANALYZE IT."
        LOCATE 11, 20
        PRINT "CHOSE 2 OR 8(END) ON MAIN MENU"
        CALL FRAME(15, 70, 8, 13)
        BEEP
        CALL waiter
        GOTO LOOP1
END IF
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
IF ((CHOICE% = 4) AND (Flag1 = 1)) THEN CALL Bargraph
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
IF ((CHOICE% = 5) AND (Flag1 = 0)) THEN
        CLS
        LOCATE 10, 20
        PRINT "YOU MUST DRAW A GRAPH BEFORE YOU CAN ANALYZE IT."
        LOCATE 11, 20
        PRINT "CHOSE 2 OR 8(END) ON MAIN MENU"
        CALL FRAME(15, 70, 8, 13)
        BEEP
        CALL waiter
        GOTO LOOP1
END IF
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
IF ((CHOICE% = 5) AND (Flag1 = 1)) THEN
        Flag3 = 1
        CALL Summarize
END IF
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
IF ((CHOICE% = 5) AND (Flag1 = 0)) THEN
        CLS
        LOCATE 10, 20
        PRINT "YOU MUST DRAW A GRAPH THEN ANALYZE IT BEFORE"
        LOCATE 11, 20
        PRINT "YOU CAN SAVE IT."
        LOCATE 12, 20
        PRINT "CHOSE 2 OR 8 (END) ON MAIN MENU"
        CALL FRAME(15, 70, 8, 13)
        BEEP
        CALL waiter
  GOTO LOOP1
  END IF
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
IF ((CHOICE% = 6) AND (Flag1 = 1) AND (Flag3 = 0)) THEN
        CLS
        LOCATE 10, 20
        PRINT "YOU MUST ANALYZE THE DATASET BEFORE YOU CAN SAVE IT"
        LOCATE 11, 20
        PRINT "DATA SET WILL BE ANALYZED NOW"
        LOCATE 12, 20
```

```
            PRINT "TO SAVE DATASET CHOSE #6 FROM THE MENU."
            CALL FRAME(15, 72, 8, 14)
            BEEP
            CALL waiter
            CALL Summarize
            Flag3 = 1
            GOTO LOOP1
      END IF
'     - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
      IF ((CHOICE% = 6) AND (Flag1 = 1) AND (Flag3 = 1)) THEN
            CALL Printfile(arraydat())
      END IF
'     - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -
      IF (CHOICE% = 7) THEN CALL Refresh
      IF CHOICE% = 8 THEN CALL Options2(anals$())
      IF CHOICE% = 9 THEN CALL Ender
GOTO LOOP1
END
'====================================================================
REM            E R R O R   H A N D L I N G   R O U T I N E S
HANDLER:

SELECT CASE ERR
      CASE 6
      BEEP
      PRINT "DIVISION BY ZERO ... CASE SKIPPED"
      LPRINT "DIVISION BY ZERO ... CASE SKIPPED"
      RESUME NEXT
      CASE 27
      BEEP
      PRINT " PRINTER OUT OF PAPER...PRESS A KEY WHEN READY"
      AAA$ = ""
      WHILE AAA$ = ""
      AAA$ = INKEY$
      WEND
      PRINT "RETRYING PRINT REQUEST..."
      RESUME

      CASE 68
      BEEP
      PRINT "PRINTER NOT READY...PUT PRINTER ON LINE AND PRESS A KEY"
      AAA$ = ""
      WHILE AAA$ = ""
      AAA$ = INKEY$
      WEND
      PRINT "RETRYING PRINT REQUEST..."
      RESUME
      CASE 25
      BEEP
      PRINT "PRINTER NOT READY...PUT PRINTER ON LINE AND PRESS A KEY"
      AAA$ = ""
      WHILE AAA$ = ""
      AAA$ = INKEY$
      WEND
      PRINT "RETRYING PRINT REQUEST..."
      RESUME
      CASE 53
            BEEP
            PRINT " FILE NOT FOUND."
            FOR x = 1 TO 150: NEXT x
```

```
                RESUME LOOP1
        CASE ELSE
        PRINT "UNANTICIPATED ERROR ENCOUNTERED CODE ", ERR
        END
END SELECT
'xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx


SUB An1
INPUT "ENTER RUN NUMBER ============>"; rn$
LPRINT CHR$(27); CHR$(15);
LPRINT CHR$(27); CHR$(9);
LPRINT "* * * *CONFIDENCE INTERVAL ANALYSIS * * * * "
LPRINT "      GARTNER LEE LIMITED"
LPRINT "SIMULATION RUN NUMBER "; rn$
LPRINT DATE$
LPRINT TIME$
LPRINT "STATISTICS CALCULATED USING POPULATION MEAN (Mu)"
timestart = TIMER
CLS : PRINT "PROGRAM RUNDATE - "; DATE$; ", RUN TIME - "; TIME$; "."
CALL READIN(arraydat()):
WIDTH LPRINT 132
FOR MONTH = 1 TO 12
IF MONTH = 7 THEN                       ' SET TO 7 FOR COMPRESSED PRINT
        BEEP                            '        13 FOR TINY PRINT
        LPRINT CHR$(12);
        aa$ = ""
        PRINT "CHANGE PAPER AND PRESS A KEY WHEN READY"
        WHILE aa$ = ""
        aa$ = INKEY$
        WEND
END IF
FLAG6 = 1

PRINT "MONTH ", MONTH
        LPRINT ""
LPRINT "MONTH ", MONTH
        LPRINT " FR   N        UCL       MEAN       LCL       RANGE       SD        VAR        AC        SKEW     EXCESS    DELTA
        LPRINT "-------------------------------------------------------------------------------------------------------------------
                LPS = 1
                CALL Sample(arraydat(), matrix(), samplenum%)
                CALL Proc1(matrix(), samplenum%)

                CALL Thrice(arraydat(), matrix(), samplenum%)
                CALL Proc1(matrix(), samplenum%)

                LPS = 7
                CALL Sample(arraydat(), matrix(), samplenum%)
                CALL Proc1(matrix(), samplenum%)
NEXT MONTH
PRINT "": PRINT "NORMAL TERMINATION AT  "; TIME$
timeend = TIMER
LPRINT ""
LPRINT "Computation time =";
LPRINT USING " ###.# "; (timeend - timestart) / 60;
LPRINT " minutes."
LPRINT CHR$(12);

END SUB
```

```
END SUB

SUB AN3
INPUT "ENTER RUN NUMBER ============>"; rn$
LPRINT CHR$(27); CHR$(15);
LPRINT CHR$(27); CHR$(9);
LPRINT "* * * *CONFIDENCE INTERVAL ANALYSIS * * * * "
LPRINT "       F U L L    A N A L Y S I S"
LPRINT "      GARTNER LEE LIMITED"
LPRINT "SIMULATION RUN NUMBER "; rn$
LPRINT DATE$
LPRINT TIME$
timestart = TIMER
CLS : PRINT "PROGRAM RUNDATE - "; DATE$; ", RUN TIME - "; TIME$; "."
CALL READIN(arraydat()):
WIDTH LPRINT 132
INPUT "ENTER MONTH TO ANALYZE (1 TO 12) "; MONTH
IF MONTH = 13 THEN                      ' SET TO 7 FOR COMPRESSED PRINT
        BEEP                            '        13 FOR TINY PRINT
        LPRINT CHR$(12);
        aa$ = ""
        PRINT "CHANGE PAPER AND PRESS A KEY WHEN READY"
        WHILE aa$ = ""
        aa$ = INKEY$
        WEND
END IF
FOR FLAG6 = 1 TO 2
        LPRINT ""
        IF FLAG6 = 1 THEN
                LPRINT "STATISTICS CALCULATED USING POPULATION MEAN (Mu) FOR MONTH "; MONTH
        END IF
        IF FLAG6 = 2 THEN
                LPRINT "STATISTICS CALCULATED USING SAMPLE MEAN (Xbar) FOR MONTH "; MONTH
        END IF
        LPRINT "---------------------------------------------------------------------------------------------------------------------------------
        LPRINT " FR   N        UCL       MEAN       LCL       RANGE       SD        VAR        AC        SKEW      EXCESS     DELTA     FRE
        LPRINT "---------------------------------------------------------------------------------------------------------------------------------
        LPRINT ""
        FOR LPS = 1 TO 15
                CALL Sample(arraydat(), matrix(), samplenum%)
                CALL Proc1(matrix(), samplenum%)
      IF LPS = 1 THEN
        LPRINT "---------------------------------------------------------------------------------------------------------------------------------
      END IF
        NEXT LPS
                CALL Thrice(arraydat(), matrix(), samplenum%)
                CALL Proc1(matrix(), samplenum%)

NEXT FLAG6
PRINT "": PRINT "NORMAL TERMINATION AT  "; TIME$
timeend = TIMER
LPRINT ""
LPRINT "Computation time =";
LPRINT USING " ###.# "; (timeend - timestart) / 60;
LPRINT " minutes."
LPRINT CHR$(12);


END SUB
```

```
SUB An2

CLS
        FOR x = 1 TO 5
        PRINT " "
NEXT x
LPRINT "* * * *PRESENCE/ABSENCE ANALYSIS PROGRAM* * * * "
LPRINT "       GARTNER LEE LIMITED"
LPRINT "       RUN DATE "; DATE$
LPRINT "       TIME "; TIME$
LPRINT " "
PRINT "                          WORKING                              "

FREQ(1) = 365    '*******SPECIFY SAMPLING INTERVALS********
FREQ(2) = 180
FREQ(3) = 90
FREQ(4) = 60
FREQ(5) = 30
FREQ(6) = 7
FREQ(7) = 2
FREQ(8) = 1
RENTER1:
        CALL readin2(arraydat()) '****READ IN DATA AND COMPUTE PRES/ABSENCE*****
LPRINT " "
LPRINT "SAMPLES HIT  MISSED  % ACTUAL    FOD       FOD-ERROR"
LPRINT "---------------------------------------------------"
FOR I = 1 TO 8
        Samplecount = INT(365 / FREQ(I))
        n = 1: NHIT = 0: Missed = 0: Hit = 0
                FOR z = 1 TO 365
                        pres = arraydat(z)
                        NHIT = 0
                        IF n MOD FREQ(I) = 0 THEN NHIT = pres
                        IF n MOD FREQ(I) <> 0 THEN NHIT = -9
                        IF pres = 1 AND NHIT = -9 THEN Missed = Missed + 1
                        IF pres = 1 AND NHIT = 1 THEN Hit = Hit + 1
                        n = n + 1
                NEXT z
        CLOSE
IF Num1 = 0 THEN Val1 = 0
IF Num1 > 0 THEN Val1 = (Hit / Num1) * 100
FOD = (Hit / Samplecount) * 100
PREERROR = FOD - (Num1 / 365) * 100
LPRINT USING " ###  "; Samplecount, Hit, Missed;
LPRINT USING "  ###.###  "; Val1; FOD; PREERROR
NEXT I

LPRINT " "
INPUT "CHANGE DETECTION LIMIT (Y/N)?"; ANS$

IF ANS$ = "y" THEN GOTO RENTER1
IF ANS$ = "Y" THEN GOTO RENTER1
GOTO ENDR
IF ANS$ = CHR$(121) THEN GOTO RENTER1
ENDR:

LPRINT CHR$(12);

LPRINT "        FINISHED"
```

```
SUB Bargraph
'====================================SCREEN BAR GRAPH ==============
CLS
SCREEN 0
pst:
INPUT "ENTER NUMBER OF CLASS INTERVALS (5 TO 20) ===> ", int.num
IF int.num > 20 OR int.num < 5 THEN GOTO pst
LOCATE 1, 65
PRINT "PLEASE WAIT...";
LOCATE 2, 20
DIM bval(20)
MIN = 999999
max = 0
CLS
FOR I = 1 TO int.num
        bval(I) = 0
NEXT I
FOR z = 1 TO 365
        xx = arraydat(z)
        IF xx > max THEN max = xx
        IF xx < MIN THEN MIN = xx
NEXT z
Range = max - MIN
int.size = CINT(Range / int.num)
PRINT "MIN             MAX           RANGE      INTERVAL SIZE"
PRINT MIN, max, Range, int.size
FOR I = 1 TO 365
        FOR a = 1 TO int.num
        temp.val = (MIN + (a - 1) * int.size)
        SELECT CASE a
                CASE IS < int.num
                IF arraydat(I) >= temp.val AND arraydat(I) < (temp.val + int.size) THEN
                        bval(a) = bval(a) + 1
                END IF
                CASE IS = int.num
                IF arraydat(I) >= temp.val AND arraydat(I) <= max THEN
                        bval(a) = bval(a) + 1
                END IF
        END SELECT

        NEXT a
NEXT I
'===========================S C R E E N S =================================
SCALE1 = 4: SCALE2 = 60: SCALE3 = 60
DSCAG:
SCREEN 0: CLS :  symbol = 177
FOR a = 1 TO int.num
        LOCATE 4 + a, 30: PRINT STRING$(bval(a) / SCALE1, symbol); bval(a);
NEXT a
LOCATE 3, 10: PRINT "       CLASS";
FOR a = 1 TO int.num
        temp.val = (MIN + (a - 1) * int.size)
        LOCATE a + 4, 10: PRINT USING "######"; temp.val;
        PRINT "   -";
        SELECT CASE a
        CASE IS < int.num
                PRINT USING "#####"; temp.val + int.size - 1
        CASE IS = int.num
                PRINT USING "#####"; max
```

```
        END SELECT
NEXT a

LOCATE 24, 10

PRINT " A B S O L U T E   F R E Q U E N C Y   D I S T R I B U T I O N";
      CALL waiter
SCREEN 0: CLS :  symbol = 177
Cumcount = 0
LOCATE 2, 60: PRINT "CUMULATIVE"
LOCATE 3, 60: PRINT "FREQUENCY "
FOR a = 1 TO int.num
        LOCATE 4 + a, 30: PRINT STRING$((bval(a) / 365 * SCALE2), symbol);
        PRINT USING "###.##"; bval(a) / 3.65;
        Cumcount = Cumcount + bval(a) / 3.65
        LOCATE 4 + a, 60: PRINT USING "###.##"; Cumcount;
NEXT a
LOCATE 3, 10: PRINT "        CLASS";
FOR a = 1 TO int.num
        temp.val = (MIN + (a - 1) * int.size)
        LOCATE a + 4, 10: PRINT USING "######"; temp.val;
        PRINT "   -";
        SELECT CASE a
        CASE IS < int.num
                PRINT USING "#####"; temp.val + int.size - 1
        CASE IS = int.num
                PRINT USING "#####"; max
        END SELECT
NEXT a

LOCATE 24, 10

PRINT " R E L A T I V E   F R E Q U E N C Y   D I S T R I B U T I O N";
      CALL waiter
SCREEN 0: CLS :  symbol = 177
Cumcount = 0
LOCATE 3, 10: PRINT "        CLASS";
LOCATE 2, 60: PRINT "CUMULATIVE": LOCATE 3, 60: PRINT "FREQUENCY"
FOR a = 1 TO int.num
        Cumcount = Cumcount + bval(a) / 3.65
        LOCATE 4 + a, 30: PRINT STRING$((Cumcount / 365 * SCALE3), symbol);
        LOCATE 4 + a, 60: PRINT USING "###.##"; Cumcount;
NEXT a
FOR a = 1 TO int.num
        temp.val = (MIN + (a - 1) * int.size)
        LOCATE a + 4, 10: PRINT USING "######"; temp.val;
        PRINT "   -";
        SELECT CASE a
        CASE IS < int.num
                PRINT USING "#####"; temp.val + int.size - 1
        CASE IS = int.num
                PRINT USING "#####"; max
        END SELECT
NEXT a

LOCATE 24, 10

PRINT " C U M U L A T I V E   F R E Q U E N C Y   D I S T R I B U T I O N
```

```
      CALL waiter
CLS : LOCATE 10, 10:
INPUT "CHANGE SCALING FACTORS? (Y/N) "; ANS$
IF ANS$ = "Y" THEN GOTO OK1
IF ANS$ = "y" THEN GOTO OK1
GOTO NOTOK
OK1:
        PRINT "SCALE1 (ABSOLUTE) SET AT "; SCALE1
        INPUT "ENTER NEW VALUE..."; SCALE1
        PRINT "SCALE2 (RELATIVE) SET AT "; SCALE2
        INPUT "ENTER NEW VALUE..."; SCALE2
        PRINT "SCALE3 (CUMULATIVE) SET AT "; SCALE3
        INPUT "ENTER NEW VALUE..."; SCALE3
GOTO DSCAG
NOTOK:
END SUB

SUB CORRELOGRAM
CLS
PRINT "CORRELOGRAM PROGRAM"
STARTTIME = TIMER
INPUT "ENTER RUN NUMBER ============>"; rn$
CALL READIN3(arraydat()):
                LPS = 1
                CALL PROC2(arraydat(), samplenum%)
LPRINT CHR$(12);
ENDTIME = TIMER
LPRINT "COMPUTATION TIME = "; (ENDTIME - STARTTIME) / 60
END SUB

SUB Draw1
CLS
Flag3 = 0
SCREEN 2
INPUT "ENTER UPPER CONCENTRATION LIMIT===>", Max.conc
INPUT "ENTER UNITS =====================>", UNITS$
INPUT "ENTER RUN # "; rn$
Yscale = (Max.conc / 84)
CLS
simlength = 365
LOCATE 2, 65
LINE (100, 180)-(500, 180)       'X-AXIS ****************************************
LINE (100, 10)-(100, 180)        'Y-AXIS ****************************************
FOR q = 1 TO 12
        LINE (100 + q * 30, 180)-(100 + q * 30, 175)
        NEXT                      'HORIZONTAL TICKS******************************
'================= L A B E L   X - A X I S =========================
                LOCATE 24, 14: PRINT 1; : LOCATE 24, 18: PRINT 2;
                LOCATE 24, 22: PRINT 3; : LOCATE 24, 25: PRINT 4;
                LOCATE 24, 29: PRINT 5; : LOCATE 24, 33: PRINT 6;
                LOCATE 24, 37: PRINT 7; : LOCATE 24, 40: PRINT 8;
                LOCATE 24, 44: PRINT 9; : LOCATE 24, 47: PRINT 10;
                LOCATE 24, 51: PRINT 11; : LOCATE 24, 55: PRINT 12;
'==================L A B E L  Y - A X I S =========================
FOR q = 1 TO 10
        LINE (95, 180 - (16.4 * q))-(465, 180 - (16.4 * q))
NEXT
```

```
LOCATE 3, 7: PRINT Max.conc
LOCATE 5, 7: PRINT Max.conc * .9
LOCATE 7, 7: PRINT Max.conc * .8
LOCATE 9, 7: PRINT Max.conc * .7
LOCATE 11, 7: PRINT Max.conc * .6
LOCATE 13, 7: PRINT Max.conc * .5
LOCATE 15, 7: PRINT Max.conc * .4
LOCATE 17, 7: PRINT Max.conc * .3
LOCATE 19, 7: PRINT Max.conc * .2
LOCATE 21, 7: PRINT Max.conc * .1
LOCATE 23, 7: PRINT 0
LOCATE 12, 2: PRINT UNITS$;
LINE (0, 0)-(600, 199), , B
LOCATE 2, 65: PRINT "RUN # "; rn$;
        BEEP                        '************************************
        AAA$ = ""                   '*  WAIT FOR USER TO PRESS A KEY    *
        WHILE AAA$ = ""             '*  BEFORE STARTING TO DRAW         *
        AAA$ = INKEY$               '*  POLLUTOGRAPH.                   *
        WEND                        '************************************
y = STICK(1)
y = 180 - (y * Yscale / Max.conc) * 164
PSET (100, y)
        IF Flagfile = 1 THEN GOTO point2
        FOR n = 2 TO simlength
                FOR JAYJAY = 1 TO DELAY: NEXT
                ystick = STICK(1): a = ystick
                correction = (((Yscale - 0 + 1) * RND + 0))
                ystick = ((ystick * Yscale) - (Yscale * 2) - (Yscale / 2)) + correction
                IF ystick < 0 THEN ystick = 0
                arraydat(n) = ystick
                'LOCATE 2, 50: PRINT USING "#######"; arraydat(n); ystick; a;
                LINE -(100 + n, 180 - (arraydat(n) / Max.conc) * 164)
        NEXT n

point2:
IF Flagfile = 1 THEN
CALL Draw2
Flagfile = 0
END IF
CALL Screensave
aa$ = ""
spec:
aa$ = INKEY$
SELECT CASE aa$
CASE "S"
        CALL Script
CASE "s"
        CALL Script
CASE "P"
        SHELL "ps"
CASE "p"
        SHELL "ps"
CASE ""
GOTO spec
CASE ELSE
END SELECT

END SUB
```

```
SUB Draw2
y = 180 - (arraydat(1) / Max.conc) * 164
PSET (100, y)
FOR n = 2 TO simlength
        LINE -(100 + n, 180 - (arraydat(n) / Max.conc) * 164)
NEXT n
Flagjk = 0'default

END SUB

SUB Ender
CLOSE
END

END SUB

SUB Fileload (arraydat())
        CLS
        CALL FRAME(6, 75, 8, 14)
        LOCATE 10, 10: PRINT "DATASET MUST BE IN FORM <C:\DESIGN\DATA00_.PRN> TO LOAD."
        LOCATE 11, 10: PRINT "IF YOU ENTER WRONG RUN # YOU WILL BE RETURNED TO MAIN MENU."
        LOCATE 12, 10: INPUT "ENTER RUN NUMBER ONLY==========>", n$
                OPEN "C:\DESIGN\DATA00" + n$ + ".PRN" FOR INPUT AS #1
        n = 1: pres.max = 0
                WHILE EOF(1) = 0
                        INPUT #1, pres
                        arraydat(n) = pres
                        IF pres > pres.max THEN pres.max = pres
                        n = n + 1
                WEND

        CLOSE
        CLS
        PRINT "MAX.CONC = "; pres.max
        CALL waiter
        Flagfile = 1
        CALL Draw1


END SUB

SUB FRAME (LEFTCOL%, RIGHTCOL%, TOPROW%, BOTTOMROW%) STATIC
        LOCATE TOPROW%, LEFTCOL%: PRINT CHR$(201)
        LOCATE TOPROW%, RIGHTCOL%: PRINT CHR$(187)
        LOCATE BOTTOMROW%, LEFTCOL%: PRINT CHR$(200);
        LOCATE BOTTOMROW%, RIGHTCOL%: PRINT CHR$(188);
                FOR VERTLINE% = TOPROW% + 1 TO BOTTOMROW% - 1
                        LOCATE VERTLINE%, LEFTCOL%: PRINT CHR$(186);
                        LOCATE VERTLINE%, RIGHTCOL%: PRINT CHR$(186);
                NEXT VERTLINE%
        HORIZLENGTH% = RIGHTCOL% - LEFTCOL% - 1
        HORIZLINE$ = STRING$(HORIZLENGTH%, 205)
        LOCATE TOPROW%, LEFTCOL% + 1: PRINT HORIZLINE$
        LOCATE BOTTOMROW%, LEFTCOL% + 1: PRINT HORIZLINE$;
END SUB

'MENU.BAS
SUB Menu (MENUCHOICES$(), NUMcHOsen%) STATIC
        CLS
        SCREEN 0
```

```
        NUMOFCHOICES% = UBOUND(MENUCHOICES$)
        PROMPT$ = " "
        OKSTRING$ = ""
        LONGSTRING% = 0
        true% = -1
        false% = 0

        FOR I% = 1 TO NUMOFCHOICES%
                FIRST$ = UCASE$(LEFT$(MENUCHOICES$(I%), 1))
                OKSTRING$ = OKSTRING$ + FIRST$
                PROMPT$ = PROMPT$ + FIRST$ + "  "
                LTEMP% = LEN(MENUCHOICES$(I%))
                IF (LTEMP% > LONGSTRING%) THEN LONGSTRING% = LTEMP%
        NEXT I%

LONGSTRING% = LONGSTRING% + 1
PROMPT$ = PROMPT$ + "->"
IF LEN(PROMPT$) >= LONGSTRING% THEN LONGSTRING% = LEN(PROMPT$) + 1
LC% = 37 - (LONGSTRING% \ 2)
RC% = 80 - LC%
TC% = 3
BC% = 10 + NUMOFCHOICES%

CALL FRAME(LC%, RC%, TC%, BC%)

FOR I% = 1 TO NUMOFCHOICES%
        LOCATE 6 + I%, LC% + 3
        PRINT UCASE$(LEFT$(MENUCHOICES$(I%), 1)) + ")" + MID$(MENUCHOICES$(I%), 2)

NEXT I%

LOCATE 4, 38: PRINT "MENU"
LINE$ = STRING$(LONGSTRING%, 196)
LOCATE 5, LC% + 3: PRINT LINE$
LOCATE 7 + NUMOFCHOICES%, LC% + 3: PRINT LINE$
LOCATE 9 + NUMOFCHOICES%, LC% + 3: PRINT PROMPT$;
CTRLKEYS$ = CHR$(13) + CHR$(27)
DONE% = false%
WHILE NOT DONE%
        LOCATE , , 1
        CHARPOS% = 0
                WHILE CHARPOS% = 0
                ANS$ = INKEY$
                IF (ANS$ <> "") THEN
                ANS$ = UCASE$(ANS$)
                CHARPOS% = INSTR(OKSTRING$, ANS$)
                IF (CHARPOS% = 0) THEN BEEP
        END IF
WEND

PRINT ANS$
LOCATE 11 + NUMOFCHOICES%, 23, 0
PRINT "<ENTER> TO CONFIRM; <ESC> TO REDO."
NUMcHOsen% = CHARPOS%
CHARPOS% = 0
WHILE CHARPOS% = 0
        ANS$ = INKEY$
        IF (ANS$ <> "") THEN
        CHARPOS% = INSTR(CTRLKEYS$, ANS$)
        IF (CHARPOS% = 0) THEN BEEP
```

```
        END IF
WEND
IF (CHARPOS% = 1) THEN
        DONE% = true%
        CLS
ELSE
        LOCATE 11 + NUMOFCHOICES%, 23: PRINT SPACE$(35)
        LOCATE 9 + NUMOFCHOICES%, LC% + 3 + LEN(PROMPT$): PRINT " ";
        LOCATE , POS(0) - 1:
        END IF
WEND
END SUB

SUB Options (OPTS$())
STARTAGAIN:
        CALL Menu(OPTS$(), CHOICE%)
        IF CHOICE% = 1 THEN
                CLS
                LOCATE 1, 20
                PRINT "PRINTER SETTINGS INITIALIZED"
                LPRINT CHR$(27); "@";
                LPRINT CHR$(7);
        END IF
        IF CHOICE% = 2 THEN
                CLS
                LOCATE 1, 20
                PRINT "TINY PRINT SET"
                LPRINT CHR$(27); "@";
                LPRINT CHR$(27); "3"; CHR$(17); : LPRINT CHR$(15); :
                LPRINT CHR$(27); "S"; "1";
                LPRINT CHR$(7);
        END IF
        IF CHOICE% = 3 THEN
        CLS
        LOCATE 10, 20
        INPUT "ENTER DRAW DELAY FACTOR (1 TO 10)...DEFAULT=1"; DELAY
        END IF
        IF CHOICE% = 4 THEN EXIT SUB
GOTO STARTAGAIN
END SUB

SUB Options2 (anals$())
STARTAGAIN2:
        CALL Menu(anals$(), CHOICE%)
        IF CHOICE% = 1 THEN CALL An1

        IF CHOICE% = 2 THEN CALL An2
        IF CHOICE% = 3 THEN CALL CORRELOGRAM
        IF CHOICE% = 4 THEN CALL AN3
        IF CHOICE% = 5 THEN EXIT SUB
GOTO STARTAGAIN2
END SUB

SUB Printfile (arraydat())
        INPUT "ENTER RUN NUMBER OR LETTER======>", a$
        B$ = "C:\design\DATA00" + a$ + ".PRN"
        OPEN B$ FOR OUTPUT AS #1
        FOR n = 1 TO simlength
        IF arraydat(n) < 0 THEN arraydat(n) = 0
        PRINT #1, arraydat(n)
```

```
            REM PRINT arraydat(n)   'remove <rem> to get screen printout
        NEXT n
        CLOSE
CLS
LOCATE 4, 18: PRINT " D E S C R I P T I V E    S T A T I S T I C S"
LOCATE 6, 20: PRINT "THE MEAN IS                         ";
PRINT USING "######.###"; xbar
LOCATE 7, 20: PRINT "THE SD IS                           ";
PRINT USING "######.###"; xsd
LOCATE 8, 20: PRINT "THE MINIMUM IS                      ";
PRINT USING "######.###"; MIN
LOCATE 9, 20: PRINT "THE MAXIMUM IS                      ";
PRINT USING "######.###"; max
LOCATE 10, 20: PRINT "RANGE IS                            ";
PRINT USING "######.###"; max - MIN
LOCATE 12, 20: PRINT "THE VARIANCE OF X IS                ";
PRINT USING "######.###"; Xvar
LOCATE 12, 20: PRINT "THE SKEWNESS COEFFICIENT IS         ";
PRINT USING "  ##.####"; skew
LOCATE 13, 20: PRINT "THE KURTOSIS COEFFICIENT IS         ";
PRINT USING "  ##.####"; Kurt
LOCATE 14, 20: PRINT "THE EXCESS COEFFICIENT IS           ";
PRINT USING "  ##.####"; excess
LOCATE 15, 20: PRINT "THE COEFFICIENT OF VARIATION IS ";
PRINT USING "  ###.#"; V
LOCATE 16, 20: PRINT "AUTOCORR. COEFFICENT (LAG=1) IS ";
PRINT USING "  ##.####"; R
CALL FRAME(15, 65, 2, 17)
LPRINT ""
LPRINT DATE$; : LPRINT " "; : LPRINT TIME$
LPRINT ""
LPRINT " D E S C R I P T I V E    S T A T I S T I C S"
LPRINT ""
LPRINT B$
LPRINT ""

LPRINT "THE MEAN IS                         ";
LPRINT USING "######.###"; xbar
LPRINT "THE SD IS                           ";
LPRINT USING "######.###"; xsd
LPRINT "THE MINIMUM IS                      ";
LPRINT USING "######.###"; MIN
LPRINT "THE MAXIMUM IS                      ";
LPRINT USING "######.###"; max
LPRINT "RANGE IS                            ";
LPRINT USING "######.###"; max - MIN
LPRINT "THE VARIANCE OF X IS                ";
LPRINT USING "######.###"; Xvar
LPRINT "THE SKEWNESS COEFFICIENT IS         ";
LPRINT USING "  ##.####"; skew
LPRINT "THE KURTOSIS COEFFICIENT IS         ";
LPRINT USING "  ##.####"; Kurt
LPRINT "THE EXCESS COEFFICIENT IS           ";
LPRINT USING "  ##.####"; excess
LPRINT "THE COEFFICIENT OF VARIATION IS ";
LPRINT USING "  ###.#"; V
LPRINT "AUTOCORR. COEFFICENT (LAG=1) IS ";
LPRINT USING "  ##.####"; R

CALL waiter
```

```
END SUB

SUB Proc1 (matrix(), samplenum%)

FOR LINES = 1 TO 3
NEXT
sumx = 0: sumy = 0: MIN = 99999: max = 0

        FOR z = 1 TO samplenum%        '  COMPUTE SUM OF X AND SUM OF Y
        sumx = sumx + matrix(z)
        IF z = 1 THEN sumy = sumy + matrix(samplenum%)
        IF z <> 1 THEN sumy = sumy + matrix(z - 1)
        NEXT z


xbar = sumx / samplenum%: ybar = sumy / samplenum%  '  COMPUTE  X AND Y MEANS
IF LPS = 1 THEN MU = xbar
IF FLAG6 = 1 THEN STANDIN = MU
IF FLAG6 = 2 THEN STANDIN = xbar
x = 0: y = 0: x2 = 0: y2 = 0: xy = 0: sumx2 = 0: sumy2 = 0: sumxy = 0
x3 = 0: x4 = 0: sumx3 = 0: sumx4 = 0
        FOR z = 1 TO samplenum%
        x = matrix(z) - STANDIN
        IF z = 1 THEN y = matrix(samplenum%) - ybar
        IF z <> 1 THEN y = matrix(z - 1) - ybar
        x2 = x ^ 2
        x3 = x ^ 3
        x4 = x ^ 4
        y2 = y ^ 2
        xy = x * y
        sumx2 = sumx2 + x2
        sumx3 = sumx3 + x3
        sumx4 = sumx4 + x4
        sumy2 = sumy2 + y2
        sumxy = sumxy + xy
        NEXT z
xsd = SQR(sumx2 / samplenum%)
IF xsd = 0 OR STANDIN = 0 THEN
        BEEP
        PRINT "ALL SAMPLES SELECTED WERE ZERO... CASE SKIPPED FOR FREQ="; LPS
        LPRINT "ALL SAMPLES SELECTED WERE ZERO... CASE SKIPPED FOR FREQ="; LPS
        EXIT SUB
END IF

ysd = SQR(sumy2 / samplenum%)
skew = sumx3 / ((xsd ^ 3) * samplenum%)
Kurt = sumx4 / ((xsd ^ 4) * samplenum%)
excess = Kurt - 3
Xvar = xsd * xsd: yvar = ysd * ysd
V = (xsd / STANDIN) * 100
IF (sumx2 * sumy2) <> 0 THEN R = sumxy / (SQR(sumx2 * sumy2))
IF (sumx2 * sumy2) = 0 THEN R = 1
COVAR = sumxy / samplenum%
V1 = samplenum% - 1
cl = t975(V1) * ((xsd) / SQR(samplenum%))
Ucl = xbar + cl
LCL = xbar - cl
IF LCL < 0 THEN LCL = 0
Range = 2 * cl
COV = (cl / STANDIN) * 100
```

```
REM PRINT USING "##"; LPS,samplenum%;
REM PRINT USING "  #####.##"; XBAR, XSD, XVAR, R, CL
SELECT CASE LPS
CASE 1
        IF samplenum% = 30 THEN LPS$ = "D"
CASE 7
        LPS$ = "W"
CASE 999
        LPS$ = "T"
CASE 2.3
        LPS$ = "T"
CASE ELSE
        LPS$ = "-"
END SELECT
LPRINT "  " + LPS$;
LPRINT USING "   ## "; samplenum%;
LPRINT USING "   ####.##"; Ucl, xbar, LCL, Range, xsd;
LPRINT USING " ######.##"; Xvar;
LPRINT USING "   ####.##"; R, skew, excess, ((xbar - MU) / MU) * 100;
LPRINT USING "     ##"; LPS
END SUB

SUB PROC2 (arraydat(), samplenum%)
OPEN "c:\DESIGN\autocor" + rn$ + ".prn" FOR OUTPUT AS #2
LPRINT "CORRELOGRAM DATA WRITTEN TO C:\DESIGN\AUTOCOR" + rn$ + "PRN"
samplenum% = 365
FOR lag = 1 TO 180
        PRINT "lag loop #"; lag
        sumx = 0: sumy = 0

                FOR z = (lag + 1) TO samplenum%      '  COMPUTE SUM OF X AND SUM OF Y
                sumx = sumx + arraydat(z)
                sumy = sumy + arraydat(z - lag)
                NEXT z

        count = samplenum% - lag
        xbar = sumx / count: ybar = sumy / count  '  COMPUTE  X AND Y MEANS
        x = 0: y = 0: x2 = 0: y2 = 0: xy = 0: sumx2 = 0: sumy2 = 0: sumxy = 0

                FOR z = (lag + 1) TO samplenum%
                x = arraydat(z) - xbar
                y = arraydat(z - lag) - ybar
                x2 = x * x
                y2 = y * y
                xy = x * y
                sumx2 = sumx2 + x2
                sumy2 = sumy2 + y2
                sumxy = sumxy + xy
                NEXT z


        IF (sumx2 * sumy2) <> 0 THEN R = sumxy / (SQR(sumx2 * sumy2))
        IF (sumx2 * sumy2) = 0 THEN R = 1
PRINT #2, lag, R
NEXT lag

END SUB

' SUBPROGRAM READIN
SUB READIN (arraydat())
```

```
        PRINT " "

        LPRINT " "
        OPEN "c:\DESIGN\DATA00" + rn$ + ".PRN" FOR INPUT AS #1
        LPRINT "DATASET " + " c:\DESIGN\DATA00" + rn$ + ".PRN"
        LPRINT ""
        n = 1
                WHILE EOF(1) = 0                'READ DATA INTO ARRAYDAT
                INPUT #1, pres
                arraydat(n) = pres
                n = n + 1
                WEND
        CLOSE
        PRINT " "
        PRINT "RAW DATA LOADED INTO 'ARRAYDAT'"
INPUT "PRINT OUT SIMULATED DATA SET ? (Y/y)", a$
IF a$ = "Y" THEN GOTO PRINT1
IF a$ = "y" THEN GOTO PRINT1
GOTO SKIP2
PRINT1:
FOR x = 1 TO 365

'LPRINT USING "      ###"; x;
LPRINT USING "######"; arraydat(x);
NEXT
BEEP
LPRINT CHR$(12);
aa$ = ""
PRINT "CHANGE PAPER AND PRESS A KEY WHEN READY"
WHILE aa$ = ""
        aa$ = INKEY$
WEND
LPRINT "SIMULATION RUN NUMBER "; rn$
LPRINT DATE$
LPRINT TIME$
LPRINT ""
SKIP2:
END SUB

SUB readin2 (arraydat())
CLS
PRINT "PRESENCE/ABSENCE ANALYSIS PROGRAM"
        INPUT "ENTER RUN NUMBER ============>"; rn$
        LPRINT "SAMPLING DATASET 'C:\DESIGN\DATA00" + rn$ + ".PRN'"
        INPUT "ENTER DETECTION LIMIT =======>"; DL
        LPRINT " "
        LPRINT "SELECTED DETECTION LIMIT ="; DL
        PRINT " "
        LPRINT " "
        OPEN "c:\DESIGN\DATA00" + rn$ + ".PRN" FOR INPUT AS #1

        REM NUM1 IS THE NUMBER OF 1'S OR NUMBER OF ABOVE DETECTION LIMIT DAYS
        n = 1: Num1 = 0
                WHILE EOF(1) = 0
                INPUT #1, pres
                temp = pres
                IF pres < DL THEN pres = 0
                IF pres >= DL THEN pres = 1
                IF pres = 1 THEN Num1 = Num1 + 1
        REM     PRINT N, temp, pres
```

```
            arraydat(n) = pres
            n = n + 1
            WEND
        CLOSE

END SUB

SUB READIN3 (arraydat())
        PRINT " "

        LPRINT " "
        OPEN "c:\DESIGN\DATA00" + rn$ + ".PRN" FOR INPUT AS #1
        LPRINT "DATASET " + " c:\DESIGN\DATA00" + rn$ + ".PRN"
        LPRINT ""
        n = 1
            WHILE EOF(1) = 0                   'READ DATA INTO ARRAYDAT
            INPUT #1, pres
            arraydat(n) = pres
            n = n + 1
            WEND
        CLOSE
        PRINT " "
        PRINT "RAW DATA LOADED INTO 'ARRAYDAT'"
INPUT "PRINT OUT SIMULATED DATA SET ? (Y/y)", a$
IF a$ = "Y" THEN GOTO PRINT2
IF a$ = "y" THEN GOTO PRINT2
GOTO SKIP3
PRINT2:
FOR x = 1 TO 365: LPRINT USING "       ###"; x; : LPRINT USING "###   "; arraydat(x); : NEXT
BEEP
LPRINT CHR$(12);
aa$ = ""
PRINT "CHANGE PAPER AND PRESS A KEY WHEN READY"
WHILE aa$ = ""
        aa$ = INKEY$
WEND
LPRINT "SIMULATION RUN NUMBER "; rn$
LPRINT DATE$
LPRINT TIME$
LPRINT ""
SKIP3:

END SUB

SUB Refresh
        SCREEN 2
        DEF SEG = &HB800
        BLOAD "C:\design\PICTURE", 0
        CALL waiter
END SUB

SUB Sample (arraydat(), matrix(), samplenum%)
        j = 0
        FOR n = ((MONTH * 30) - 29) TO (MONTH * 30)
         IF n MOD LPS = 0 THEN
            j = j + 1
            matrix(j) = arraydat(n)
        END IF
        samplenum% = j
        CLOSE
```

```
        NEXT
END SUB

SUB Screensave
'LOCATE 1, 65: PRINT "                  "
'LOCATE 2, 65: PRINT "                  "
DEF SEG = &HB800
BSAVE "C:\design\PICTURE", 0, &H4000
END SUB

SUB Script
        B$ = "C:\design\DATA00" + rn$ + ".PRN"
        OPEN B$ FOR OUTPUT AS #1
        FOR n = 1 TO simlength
        IF arraydat(n) < 0 THEN arraydat(n) = 0
        PRINT #1, arraydat(n)
                REM PRINT arraydat(n)   'remove <rem> to get screen printout
        NEXT n
        CLOSE
        LPRINT "DATA SET WRITTEN TO "; B$
END SUB

SUB Summarize
CLS
SCREEN 0
LOCATE 1, 65
PRINT "PLEASE WAIT..."
sumx = 0: sumy = 0: MIN = 99999: max = 0

        FOR z = 1 TO 365        '  COMPUTE SUM OF X AND SUM OF Y
                sumx = sumx + arraydat(z)
                IF arraydat(z) < MIN THEN MIN = arraydat(z)
                IF arraydat(z) > max THEN max = arraydat(z)
                IF z > 1 THEN sumy = sumy + arraydat(z - 1)
        NEXT z

xbar = sumx / 365: ybar = sumy / 365  '  COMPUTE  X AND Y MEANS

x = 0: y = 0: x2 = 0: x3 = 0: x4 = 0: y2 = 0: xy = 0:
sumx2 = 0: sumx3 = 0: sumx4 = 0: sumy2 = 0: sumxy = 0: sumx3 = 0

        FOR z = 1 TO 365
        x = arraydat(z) - xbar
        IF z > 1 THEN y = arraydat(z - 1) - ybar
        x2 = x ^ 2
        x3 = x ^ 3
        x4 = x ^ 4
        y2 = y * y
        xy = x * y
        sumx2 = sumx2 + x2
        sumx3 = sumx3 + x3
        sumx4 = sumx4 + x4
        sumy2 = sumy2 + y2
        sumxy = sumxy + xy
        NEXT z

xsd = SQR(sumx2 / 365)
ysd = SQR(sumy2 / 365)
skew = sumx3 / ((xsd ^ 3) * 365) 'FOR N=365 ONLY!
Kurt = sumx4 / ((xsd ^ 4) * 365)
```

```
Xvar = xsd * xsd: yvar = ysd * ysd
V = (xsd / xbar) * 100
excess = Kurt - 3
R = sumxy / (SQR(sumx2 * sumy2))
CLS
LOCATE 4, 18: PRINT " D E S C R I P T I V E    S T A T I S T I C S"
LOCATE 6, 20: PRINT "THE MEAN IS                         ";
PRINT USING "######.###"; xbar
LOCATE 7, 20: PRINT "THE SD IS                           ";
PRINT USING "######.###"; xsd
LOCATE 8, 20: PRINT "THE MINIMUM IS                      ";
PRINT USING "######.###"; MIN
LOCATE 9, 20: PRINT "THE MAXIMUM IS                      ";
PRINT USING "######.###"; max
LOCATE 10, 20: PRINT "RANGE IS                            ";
PRINT USING "######.###"; max - MIN
LOCATE 11, 20: PRINT "THE VARIANCE OF X IS                ";
PRINT USING "######.###"; Xvar
LOCATE 12, 20: PRINT "THE SKEWNESS COEFFICIENT IS         ";
PRINT USING "  ##.####"; skew
LOCATE 13, 20: PRINT "THE KURTOSIS COEFFICIENT IS         ";
PRINT USING "  ##.####"; Kurt
LOCATE 14, 20: PRINT "THE EXCESS COEFFICIENT IS           ";
PRINT USING "  ##.####"; excess
LOCATE 15, 20: PRINT "THE COEFFICIENT OF VARIATION IS ";
PRINT USING "  ###.#"; V
LOCATE 16, 20: PRINT "AUTOCORR. COEFFICENT (LAG=1) IS ";
PRINT USING "  ##.####"; R
CALL FRAME(15, 65, 2, 17)
CALL waiter

END SUB

SUB Thrice (arraydat(), matrix(), samplenum%)
        LPS = 999                        'RESET SO THAT MU IS NOT RECALCULATED
        FLAG = -1
        j = 0
        FOR n = ((MONTH * 30) - 29) TO (MONTH * 30)
                IF n MOD 7 = 0 THEN
                        FLAG = FLAG * -1
                        GOTO SKIP
                END IF
                 SELECT CASE FLAG
                        CASE -1
                        IF n MOD 2 = 0 THEN
                                j = j + 1
                                matrix(j) = arraydat(n)
             REM           LPRINT N, matrix(j)
                        END IF
                        CASE 1
                        IF (n - 1) MOD 2 = 0 THEN
                                j = j + 1
                                matrix(j) = arraydat(n)
             REM          LPRINT N, matrix(j)
                        END IF
                 END SELECT
        IF j = 0 THEN GOTO SKIP

SKIP:
```

```
samplenum% = j
        NEXT
        LPS = 2.3
END SUB

SUB waiter
LOCATE 25, 65
PRINT "PRESS A KEY...";
aa$ = INKEY$
        WHILE aa$ = ""
        aa$ = INKEY$
        WEND
END SUB
```

```
#
# This macro generates a binary (0/1) series (e.g., a time series) having
#   specified parameters. Then a macro ('BINSER3.DAT') can be executed to
#   to analyze the series and to evaluate the efficiency of various
#   sampling schemes. For more detail, leave MINITAB and read the file
#   'BINSER.DOC'.
#
# Enter "LET K1 = the mean mu of the series, i.e. the true long-term mean
#   probability of state 1 occurring"
# Enter "LET K2 = the autocorrelation coefficient rho (lag 1)"
# Enter "LET K3 = the length N of the series to be generated"
#
# Enter "EXEC 'BINSER1.DAT' "
```

```
# The true long-term mean probability of state 1 is:
PRINT K1
# The autocorrelation coefficient is:
PRINT K2
# The length of the series to be generated is:
PRINT K3
LET K8=K3-1
LET K6=K8*K2*K2
LET K7=1-K1
#
# The expected chi-square (1 df) is:
PRINT K6
LET K9=1-K1+K2*K1
LET K10=1-K9
SET C4
 0 1
END
SET C5
 K9 K10
END
OH=0
LET K10=K1+K2*(1-K1)
LET K9=1-K10
SET C6
 0 1
END
SET C7
 K9 K10
END
SET C8
 0 1
END
SET C9
 K7 K1
END
RANDOM 1 C10;
 DISCRETE C8 C9.
LET K9=1
EXEC 'BINSER2.DAT' K8 TIMES
OH=24
#
# The series has been generated, and is in column C10. If you want
#   to analyze it, enter "EXEC 'BINSER3.DAT' ".
#
```

```
LET K10=4+2*C10(K9)
LET K11=K10+1
RANDOM 1 C11:
 DISCRETE CK10 CK11.
STACK C10 C11,C10
LET K9=K9+1
```

```
#
# You may be executing this macro as a follow-up to 'BINSER1.DAT'
#   which would have been used to generate a binary series with
#   specified parameters and to store it in column C10. If so, you will
#   interpret the following analysis and evaluation as a description of a
#   simulated series which has known (specified) parameters. Thus the
#   sample estimates of the parameters (from 'BINSER3.DAT') can be
#   compared to their known values (the input to 'BINSER1.DAT').
#
# Alternatively, the binary series (which must be stored in column C10)
#   could be a real monitoring data set where state 1 represents
#   "detectable" or "violation of the standard" and state 0 represents
#   "below detectable" or "not in violation of the standard". If this is
#   the case, then the following analysis and evaluation is for data
#   whose parameters are not known but for which estimates are desired.
#   You may wish to use these parameter estimates (of mu and rho )
#   as input to 'BINSER1.DAT', to simulate binary series having these
#   properties, and then to run 'BINSER3.DAT' again to evaluate the
#   efficiency of various sampling schemes for use in sampling such
#   monitoring data.
#
# Now we will analyze the binary series:
PRINT C10
TSPLOT C10
MEAN C10
RUNS 0.5 C10
LET C13(1)=C10(1)
LET K12=1
LET K13=2
LET K3=COUNT(C10)
LET K8=K3-1
OH=0
EXEC 'BINSER4.DAT' K8 TIMES
OH=24
CODE (-1)0 C13,C13
PARSUM C13,C14
LET C15=C10*C14
LET K12=MAX1(C15)
#
# The number of runs of 1 is:
PRINT K12
#
# Here is the binary series again, with the 1's replaced by the run
#   number (1st, 2nd, --- run of 1's):
PRINT C15
ACF 5 C10
COPY C10 C12;
 OMIT 1:1.
COPY C10 C11;
 OMIT K3:K3.
CORR C11 C12,M1
COPY M1 C12-C13
LET K9=C13(1)
LET K10=K8*K9*K9
# The autocorrelation coefficient and the chi-square (1 df) are:
```

```
PRINT K9 K10
#
# Now we only observe at intervals (e.g, the series is daily and you
#   are sampling every so many days).
#
# LET K14 = The sampling (observation) interval you want, e.g. 3 if you
#   want every 3 days, and then enter "EXEC 'BINSER5.DAT' ".
```

```
LET C13(K13)=C10(K13)-C10(K12)
LET K12=K12+1
LET K13=K13+1
```

```
#
# The sampling interval is now:
PRINT K14
LET K9=ROUND(K3/K14+0.5)
LET K15=K14-1
SET C12
K15(0)
END
STACK C12 1,C12
OH=0
EXEC 'BINSER6.DAT' K9 TIMES
OH=24
COPY C11 C11;
 USE 1:K3.
LET C12=C15*C11
NAME C15 'SERIES',C11 'OBSERVE',C12 'DETECTED'
#
# Following are the generated series (col.1), observation times (col.2),
#   and occurrences of 1 that are detected (col.3):
PRINT C15 C11 C12
LET C16=C11+C15
#
# The number of runs of 1 that were missed by a given sampling scheme can
#   be calculated as the number of runs of 1 minus the number of non-zero
#   categories listed in the following TALLYs of the data.
TALLY C16
#
# You can change the value stored in K14 and "EXEC 'BINSER5.DAT'" again.
```

```
STACK C12 C11,C11
```

APPENDIX C

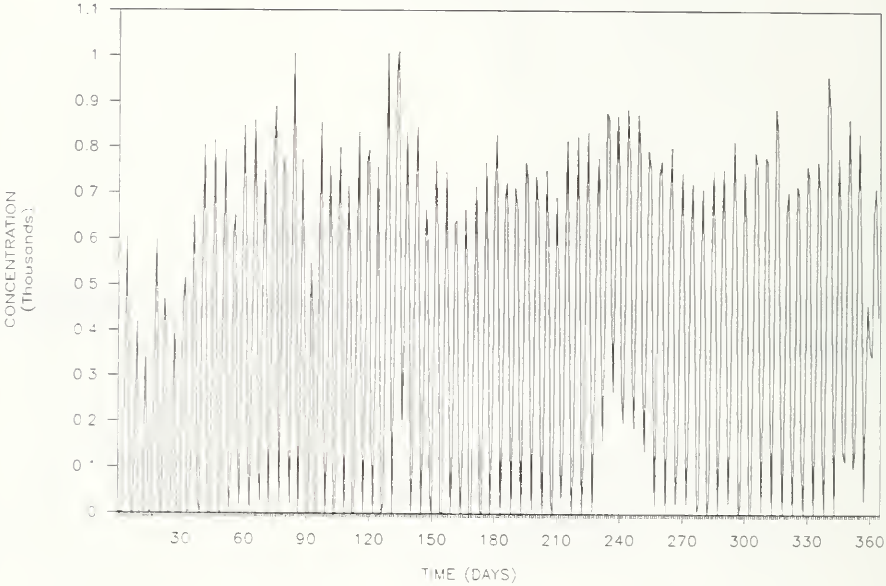
SIMULATED DATA SETS USED AS

EXAMPLES IN THE REPORT

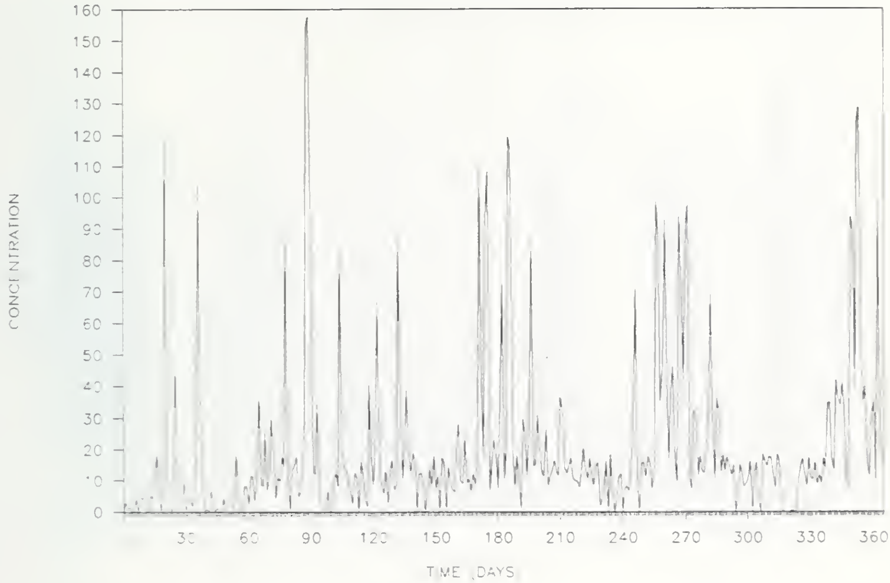
C1: TIME SERIES SCATTERGRAMS

Following are plots of concentration vs. time for the 19 simulated data sets used as examples in the report.

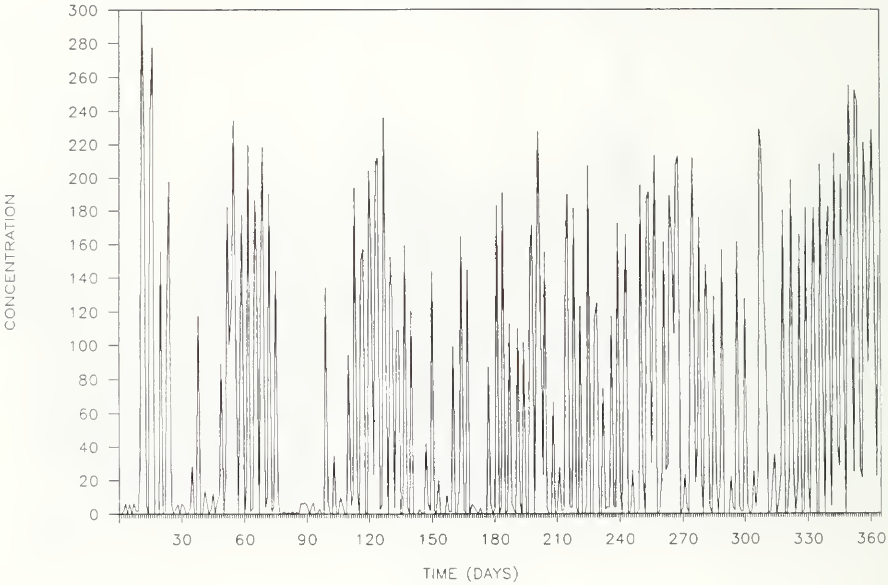
TIME SERIES SCATTERPLOT FOR RUN 01



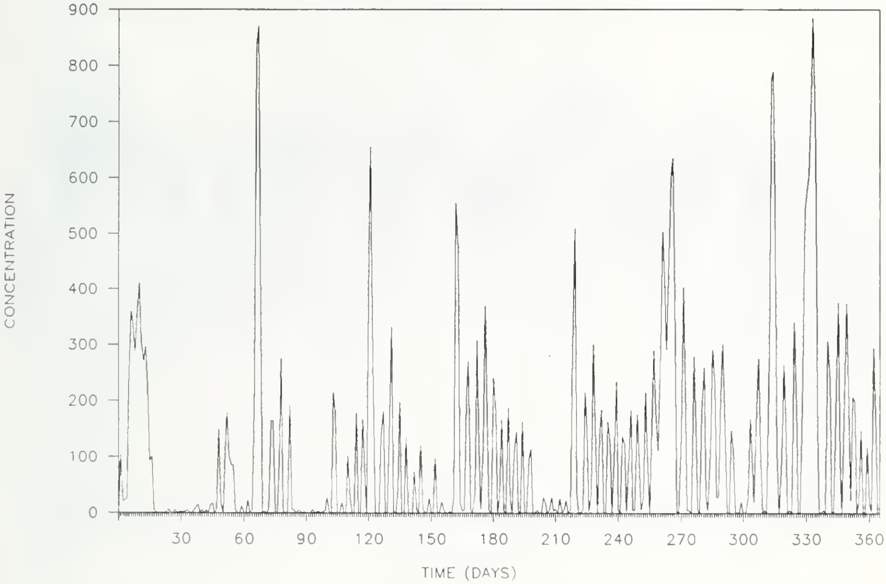
TIME SERIES SCATTERPLOT FOR RUN 02



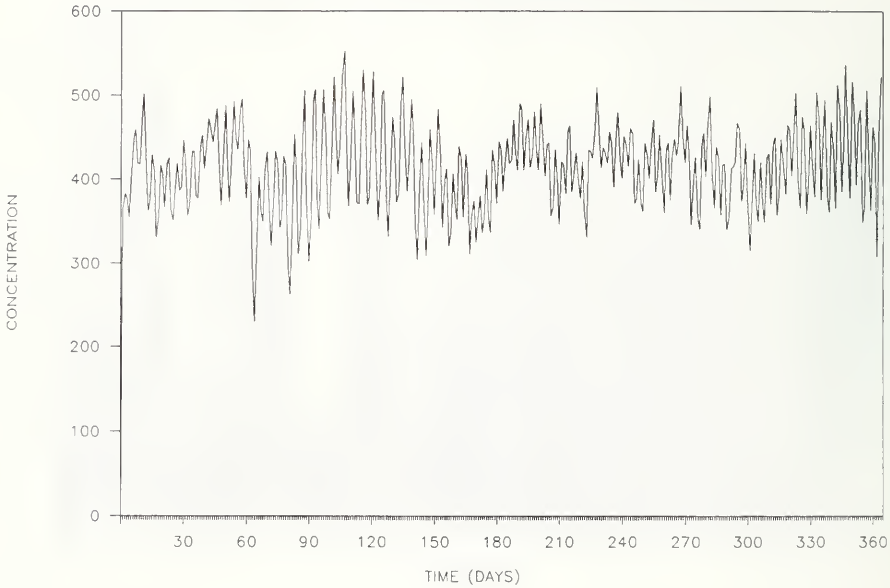
TIME SERIES SCATTERPLOT FOR RUN 03



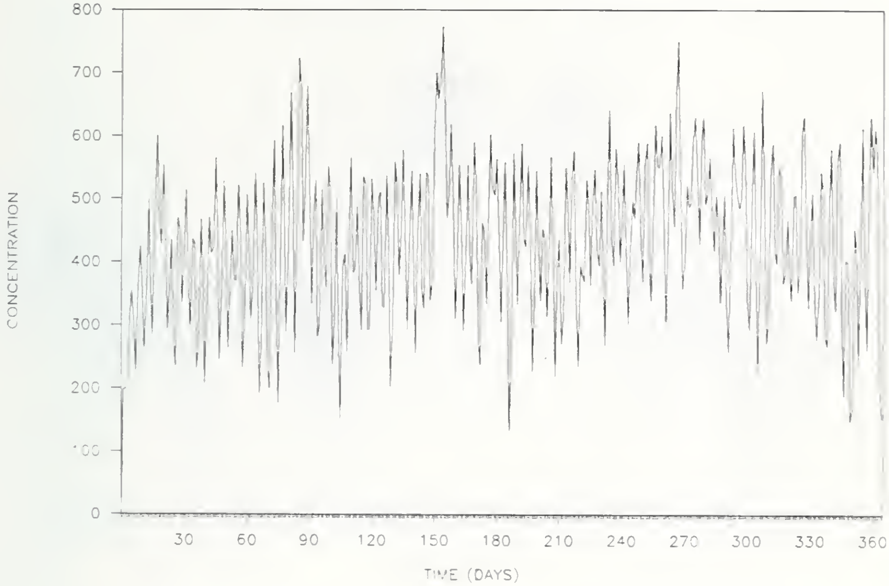
TIME SERIES SCATTERPLOT FOR RUN 04



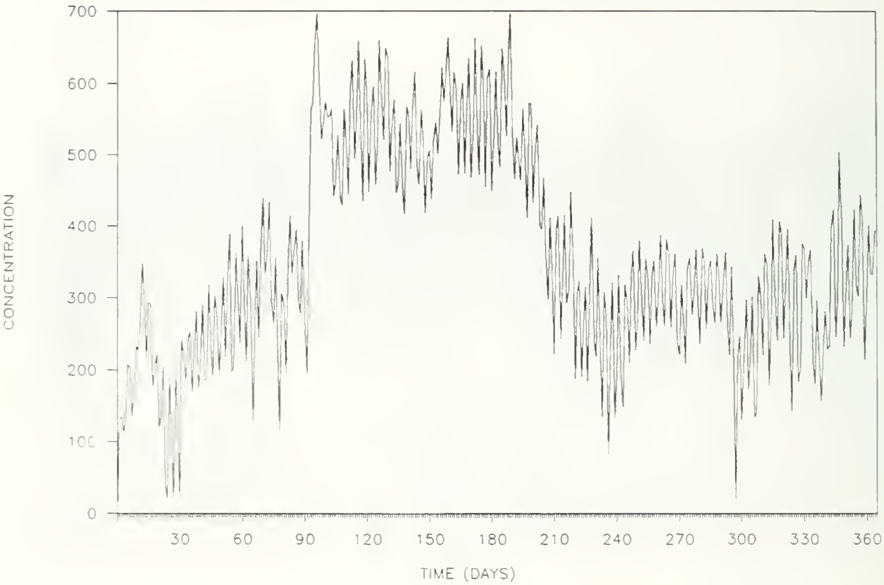
TIME SERIES SCATTERPLOT FOR RUN 05



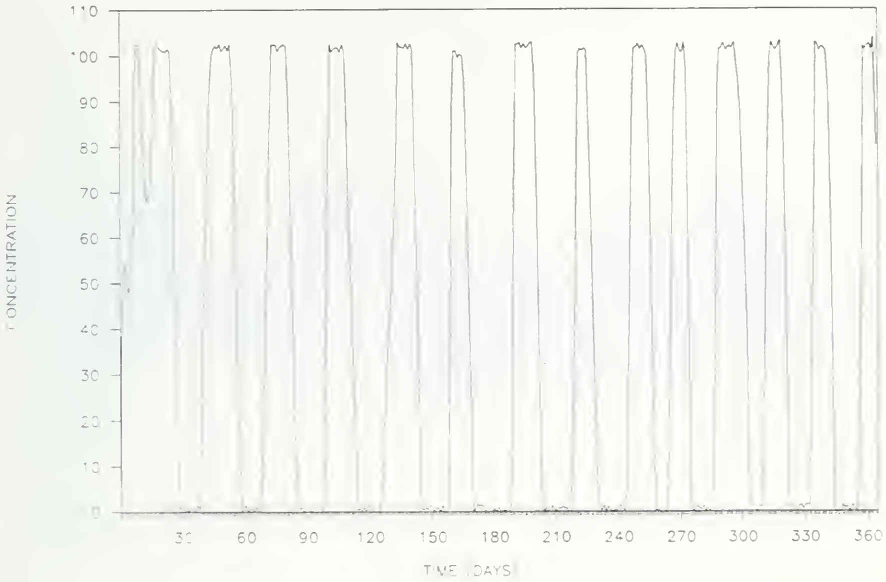
TIME SERIES SCATTERPLOT FOR RUN 06



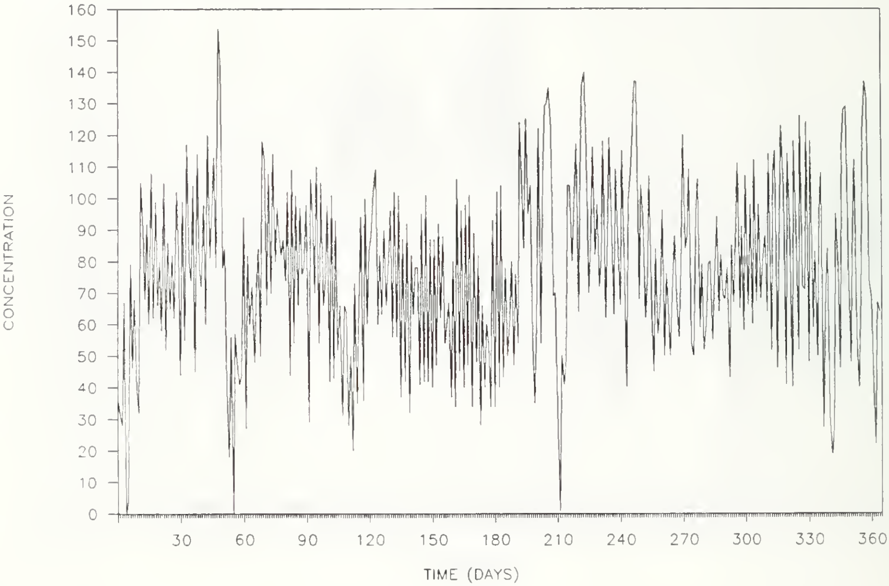
TIME SERIES SCATTERPLOT FOR RUN 07



TIME SERIES SCATTERPLOT FOR RUN 08

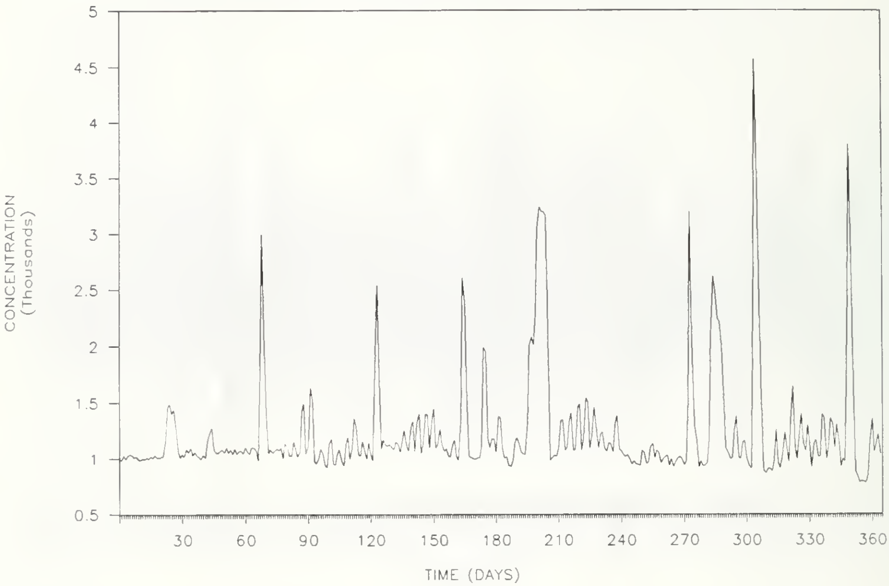


TIME SERIES SCATTERPLOT FOR RUN 09

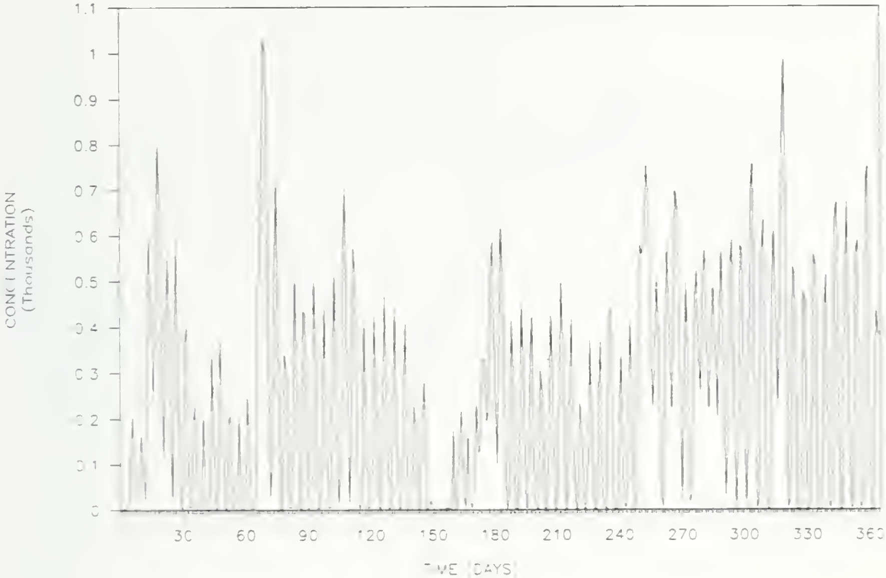



TIME SERIES SCATTERPLOT FOR RUN 10
CONCENTRATION
(Thousands)
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
30
60
90
120
150
180
210
240
270
300
330
360
TIME (DAYS)

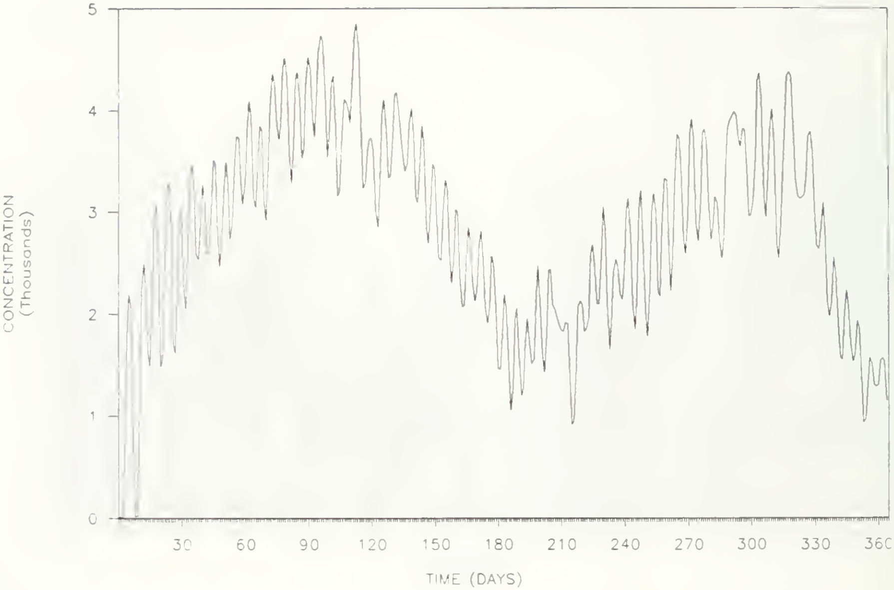
TIME SERIES SCATTERPLOT FOR RUN 11



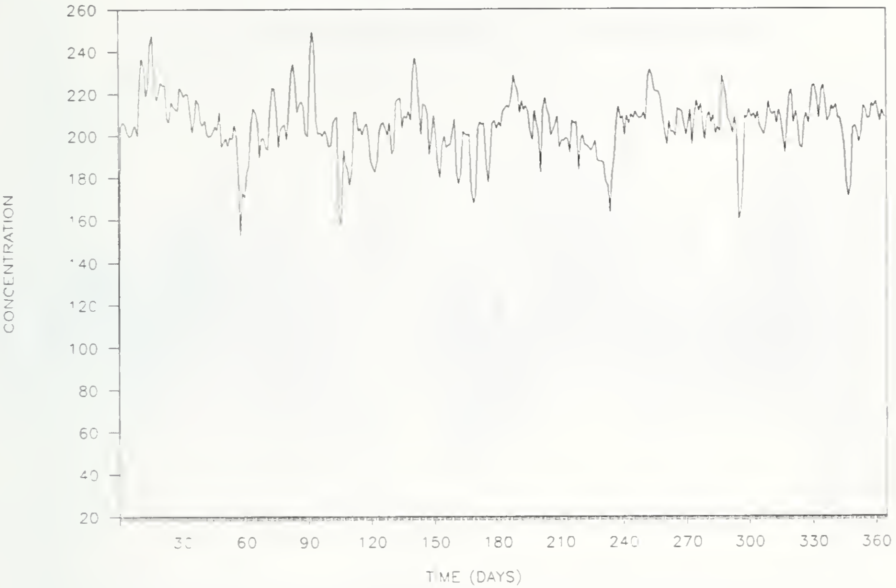
TIME SERIES SCATTERPLOT FOR RUN 12



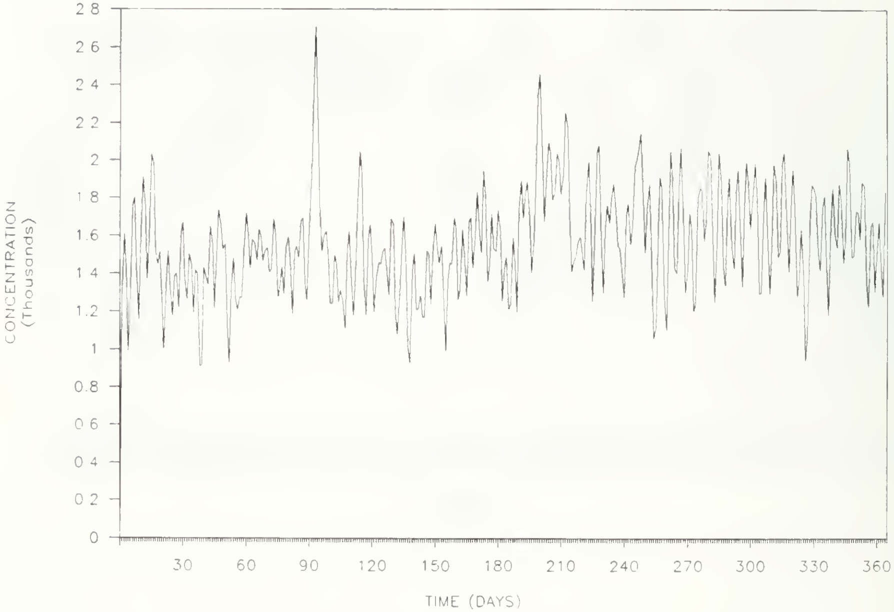
TIME SERIES SCATTERPLOT FOR RUN 13



TIME SERIES SCATTERPLOT FOR RUN 14

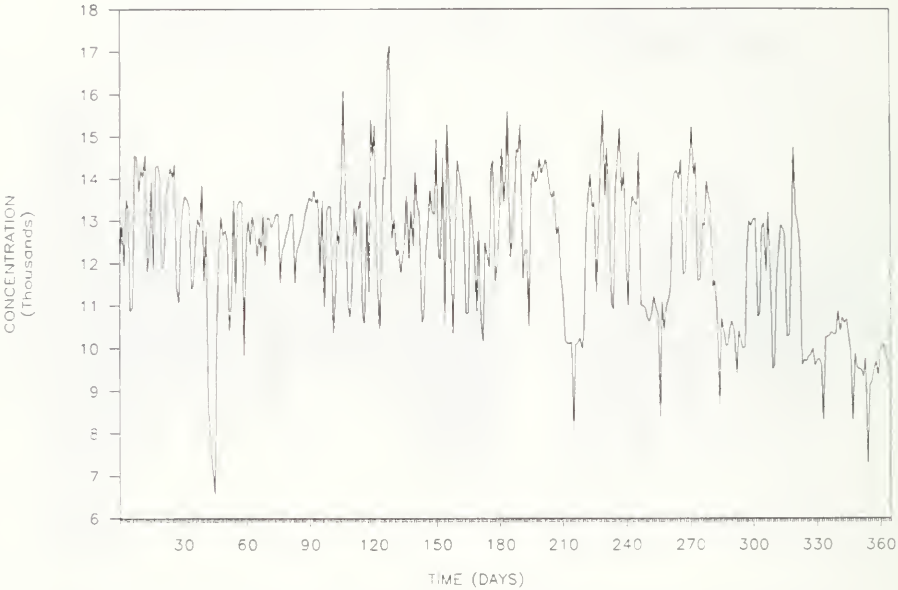


TIME SERIES SCATTERPLOT FOR RUN 15

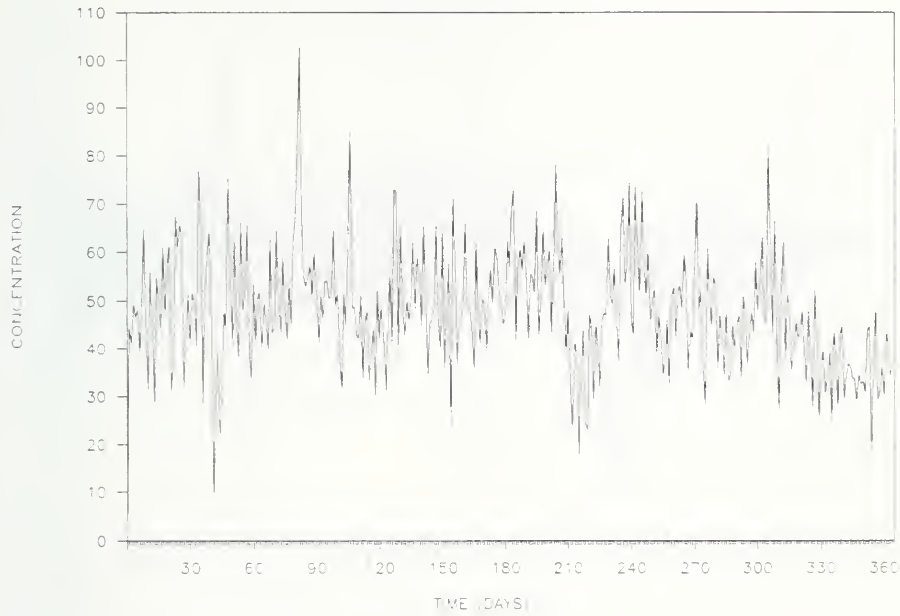



TIME SERIES SCATTERPLOT FOR RUN 16
CONCENTRATION
TIME (DAYS)
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
30
60
90
120
150
180
210
240
270
300
330
360

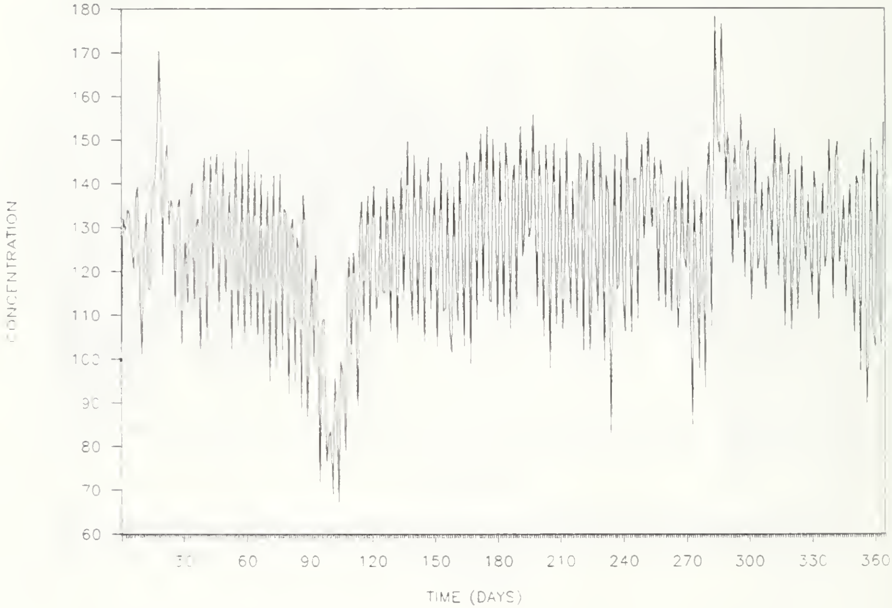
TIME SERIES SCATTERPLOT FOR RUN 17



TIME SERIES SCATTERPLOT FOR RUN 18



TIME SERIES SCATTERPLOT FOR RUN 19



C2: DESCRIPTIVE STATISTICS

Following are basic descriptive statistics for the 19 simulated data sets used in the report. The statistics were calculated using STATPAC. The data sets are contained on the distribution disk.

SUMMARY OF SIMULATED DATASETS.

RUN #	KOLMOGOROV-SMIRNOV STATISTIC
1	2.5434
2	5.4244
3	5.6324
4	5.0121
5	0.9643
6	1.0545
7	1.4440
8	4.2958
9	0.7927
10	0.8333
11	5.1572
12	3.8969
13	1.0032
14	2.5905
15	3.8268
16	0.8071
17	1.5542
18	1.1431
19	6.3586

The Kolmorgorov-Smirnov statistic provides a quick check to determine the degree of normality in a dataset. The value provides a relative indication of normality; as the value moves further from zero we can be more certain that the data do not approximate a normal distribution. The distribution is non-normal at the .025 level if KS>.955.

Using this criterion run numbers 5, 9, 10 and 16 are considered to be approximately normally distributed.

StatPac Gold Statistical Analysis Package

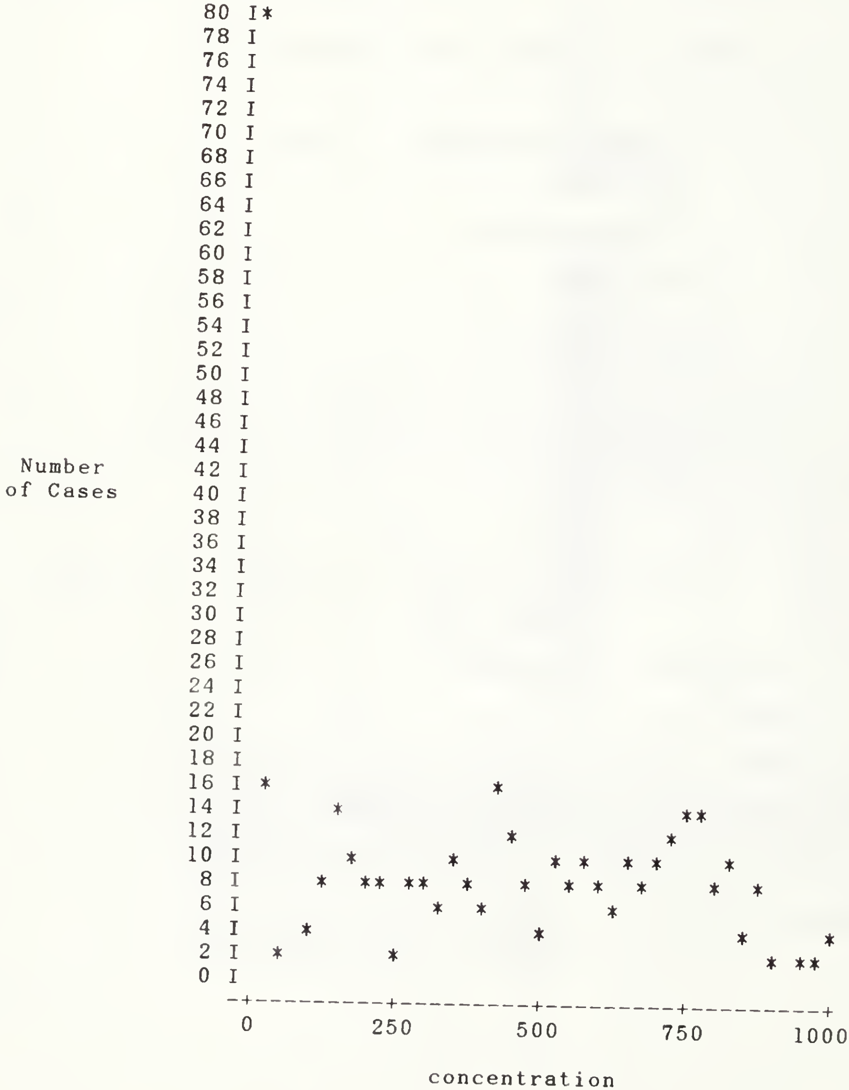
DESCRIPTIVE STATISTICS FOR RUN 01

concentration

```
     Minimum                  =  0

     Maximum                  =  1010.56

     Range                    =  1010.5600

     Sum                      =  139552.3106092

     Mean                     =  382.3351

     Median                   =  380.7280

     Mode                     =  0

     Variance                 =  91134.3712

     Standard deviation       =  301.8847

     Standard error of the mean  =  15.8231

     95 Percent confidence interval around the mean =  351.3219 - 413.3483

     Variance (unbiased)              =  91384.7403

     Standard deviation (unbiased)    =  302.2991

     Skewness        =  0.1249

     Kurtosis        =  1.6623

     Kolmogorov-Smirnov statistic for normality =  2.5434

Valid cases      = 365
Missing cases    = 0
Response percent = 100.0 %
```

StatPac Gold Statistical Analysis Package

DESCRIPTIVE STATISTICS FOR RUN 01



tatPac Gold Statistical Analysis Package

ESCRIPTIVE STATISTICS FOR RUN 02

oncentration

Minimum	=	0
Maximum	=	157.5583
Range	=	157.5583
Sum	=	7395.5723875
Mean	=	20.2618
Median	=	12.1780
Mode	=	0
Variance	=	706.0604
Standard deviation	=	26.5718

Standard error of the mean = 1.3927

95 Percent confidence interval around the mean = 17.5321 - 22.9916

Variance (unbiased) = 708.0001

Standard deviation (unbiased) = 26.6083

Skewness = 2.6104

Kurtosis = 10.0423

Kolmogorov-Smirnov statistic for normality = 5.4244

alid cases = 365
issing cases = 0
esponse percent = 100.0 %

StatPac Gold Statistical Analysis Package

DESCRIPTIVE STATISTICS FOR RUN 02

```
              88 I  *
              86 I
              84 I
              82 I
              80 I
              78 I
              76 I
              74 I
              72 I
              70 I
              68 I
              66 I
              64 I   *
              62 I
              60 I
              58 I*
              56 I
              54 I
              52 I
              50 I
              48 I
  Number      46 I *
 of Cases     44 I
              42 I
              40 I
              38 I
              36 I
              34 I
              32 I    *
              30 I
              28 I
              26 I
              24 I
              22 I
              20 I
              18 I       *
              16 I
              14 I        *
              12 I
              10 I      *
               8 I
               6 I     *             *
               4 I                 ** *
               2 I         **  ***     ******    **
               0 I
                 -+---------+---------+---------+--
                  0        50        100       150

                           concentration
```

StatPac Gold Statistical Analysis Package

DESCRIPTIVE STATISTICS FOR RUN 03

concentration

```
     Minimum                   =  0

     Maximum                   =  299.0852

     Range                     =  299.0852

     Sum                       =  20758.9140527

     Mean                      =  56.8737

     Median                    =  6.2462

     Mode                      =  0

     Variance                  =  6122.0974

     Standard deviation        =  78.2438

     Standard error of the mean  =  4.1011

     95 Percent confidence interval around the mean =  48.8356 - 64.9119

     Variance (unbiased)               =  6138.9164

     Standard deviation (unbiased)     =  78.3512

     Skewness         =  1.0821

     Kurtosis         =  2.6903

     Kolmogorov-Smirnov statistic for normality =  5.6324

Valid cases       = 365
Missing cases     = 0
Response percent  = 100.0 %
```

StatPac Gold Statistical Analysis Package

DESCRIPTIVE STATISTICS FOR RUN 03

```
            140 I*
            135 I
            130 I
            125 I
            120 I
            115 I
            110 I
            105 I
            100 I
             95 I
             90 I
             85 I
             80 I
 Number      75 I
of Cases     70 I
             65 I
             60 I *
             55 I
             50 I
             45 I
             40 I
             35 I
             30 I
             25 I
             20 I     *
             15 I
             10 I    *
              5 I  *   * *         *** ** ** ****** ***** **** *
              0 I   *   *     * ***   *  *  *      *     *    * *****    *   *
                -+---------+---------+---------+---------+---------+---------+
                 0        50        100       150       200       250       30

                                        concentration
```

StatPac Gold Statistical Analysis Package

DESCRIPTIVE STATISTICS FOR RUN 04

concentration

Minimum = 0

Maximum = 885.0707

Range = 885.0707

Sum = 39388.6643124

Mean = 107.9141

Median = 15.3875

Mode = 0

Variance = 27753.8737

Standard deviation = 166.5949

Standard error of the mean = 8.7319

95 Percent confidence interval around the mean = 90.7995 - 125.0288

Variance (unbiased) = 27830.1206

Standard deviation (unbiased) = 166.8236

Skewness = 2.1538

Kurtosis = 8.1432

Kolmogorov-Smirnov statistic for normality = 5.0121

Valid cases = 365
Missing cases = 0
Response percent = 100.0 %

StatPac Gold Statistical Analysis Package

DESCRIPTIVE STATISTICS FOR RUN 04

```
            180 I*
            175 I
            170 I
            165 I
            160 I
            155 I
            150 I
            145 I
            140 I
            135 I
            130 I
            125 I
            120 I
            115 I
            110 I
            105 I
            100 I
 Number      95 I
of Cases     90 I
             85 I
             80 I
             75 I
             70 I
             65 I
             60 I
             55 I
             50 I
             45 I
             40 I
             35 I *
             30 I
             25 I
             20 I    *  *
             15 I     *      *
             10 I      * ** *
              5 I   *      *    **  *
              0 I  *          **  *  * *****   **** *
                -+---------+---------+---------+-----
                 0        250       500       750

                            concentration
```

StatPac Gold Statistical Analysis Package

DESCRIPTIVE STATISTICS FOR RUN 05

concentration

```
    Minimum                   =  229.651

    Maximum                   =  552.093

    Range                     =  322.4420

    Sum                       =  151210.1404

    Mean                      =  414.2744

    Median                    =  418.4287

    Mode                      =  Multi-Modal

    Variance                  =  2704.5700

    Standard deviation        =  52.0055

    Standard error of the mean  =  2.7258

    95 Percent confidence interval around the mean =  408.9317 - 419.6170

    Variance (unbiased)               =  2712.0001

    Standard deviation (unbiased)     =  52.0769

    Skewness        = -0.1060

    Kurtosis        =  2.9156

    Kolmogorov-Smirnov statistic for normality =  0.9643

Valid cases      = 365
Missing cases    = 0
Response percent = 100.0 %
```

StatPac Gold Statistical Analysis Package

DESCRIPTIVE STATISTICS FOR RUN 05

```
            37 I                    *
            36 I
            35 I
            34 I
            33 I
            32 I
            31 I
            30 I                     *
            29 I                      *
            28 I                       *
            27 I               *
            26 I
            25 I                *
            24 I
            23 I              *
            22 I                   *
            21 I
  Number    20 I
of Cases    19 I
            18 I                        *
            17 I
            16 I
            15 I             *
            14 I                  *
            13 I                 *
            12 I                         *
            11 I                          *
            10 I
             9 I            *
             8 I                            **
             7 I                           *
             6 I           *
             5 I        *                     *
             4 I         *
             3 I          *
             2 I                               *
             1 I *  *                           **
               ---------+---------+---------+-----
                       300       400       500

                         concentration
```

StatPac Gold Statistical Analysis Package

DESCRIPTIVE STATISTICS FOR RUN 06

concentration

Minimum	=	118.2183
Maximum	=	1900
Range	=	1781.7817
Sum	=	159334.1992
Mean	=	436.5321
Median	=	435.2805
Mode	=	Multi-Modal
Variance	=	21042.6954
Standard deviation	=	145.0610

Standard error of the mean = 7.6033

95 Percent confidence interval around the mean = 421.6297 - 451.4344

Variance (unbiased) = 21100.5050

Standard deviation (unbiased) = 145.2601

Skewness = 2.7177

Kurtosis = 29.6985

Kolmogorov-Smirnov statistic for normality = 1.0545

Valid cases = 365
Missing cases = 0
Response percent = 100.0 %

StatPac Gold Statistical Analysis Package

DESCRIPTIVE STATISTICS FOR RUN 06

```
             54 I       *
             52 I        *
             50 I
             48 I     *
             46 I      *  *
             44 I
             42 I
             40 I
             38 I
             36 I
             34 I
             32 I    *
             30 I
 Number      28 I   *      *
of Cases     26 I
             24 I
             22 I
             20 I
             18 I
             16 I
             14 I
             12 I  *        *
             10 I
              8 I
              6 I            *
              4 I *
              2 I*            *                      *
              0 I
                ---------+---------+---------+--------
                        500      1000      1500

                            concentration
```

StatPac Gold Statistical Analysis Package

DESCRIPTIVE STATISTICS FOR RUN 07

concentration

Minimum	=	17.90778
Maximum	=	1300
Range	=	1282.0922
Sum	=	131298.80505
Mean	=	359.7228
Median	=	333.3117
Mode	=	Multi-Modal
Variance	=	24350.9750
Standard deviation	=	156.0480
Standard error of the mean	=	8.1791
95 Percent confidence interval around the mean	=	343.6917 - 375.7539
Variance (unbiased)	=	24417.8733
Standard deviation (unbiased)	=	156.2622
Skewness	=	0.7851
Kurtosis	=	5.4773
Kolmogorov-Smirnov statistic for normality	=	1.4440

Valid cases = 365
Missing cases = 0
Response percent = 100.0 %

StatPac Gold Statistical Analysis Package

DESCRIPTIVE STATISTICS FOR RUN 07

```
             34 I            *
             33 I
             32 I
             31 I
             30 I
             29 I
             28 I
             27 I         *
             26 I              *
             25 I             *
             24 I
             23 I        * *
             22 I
             21 I
             20 I
             19 I                *
  Number     18 I
of Cases     17 I
             16 I           *      *
             15 I                     * *
             14 I     *
             13 I               *
             12 I       *         *    *
             11 I                   *
             10 I                        **
              9 I                    *
              8 I
              7 I                          *
              6 I
              5 I
              4 I *    *
              3 I
              2 I                           **
              1 I  ***                                               *
                -+---------+---------+---------+---------+---------+--
                 0        250       500       750      1000      1250

                                    concentration
```

StatPac Gold Statistical Analysis Package

DESCRIPTIVE STATISTICS FOR RUN 08

concentration

Minimum	=	0
Maximum	=	440
Range	=	440
Sum	=	16885.4836029
Mean	=	46.2616
Median	=	36.6346
Mode	=	0
Variance	=	2384.2627
Standard deviation	=	48.8289
Standard error of the mean	=	2.5593
95 Percent confidence interval around the mean	=	41.2453 - 51.2779
Variance (unbiased)	=	2390.8129
Standard deviation (unbiased)	=	48.8959
Skewness	=	1.5427
Kurtosis	=	12.4241
Kolmogorov-Smirnov statistic for normality	=	4.2958

Valid cases = 365
Missing cases = 0
Response percent = 100.0 %

StatPac Gold Statistical Analysis Package

DESCRIPTIVE STATISTICS FOR RUN 08

```
            155 I*
            150 I
            145 I
            140 I
            135 I
            130 I
            125 I
            120 I
            115 I
            110 I
            105 I          *
            100 I
             95 I
             90 I
             85 I
 Number      80 I
of Cases     75 I
             70 I
             65 I
             60 I
             55 I
             50 I
             45 I
             40 I
             35 I
             30 I
             25 I
             20 I
             15 I    *  * *
             10 I ***  ** *
              5 I
              0 I                                            *
                -+---------+---------+---------+---------+----
                 0        100       200       300       400

                                concentration
```

StatPac Gold Statistical Analysis Package

DESCRIPTIVE STATISTICS FOR RUN 09

concentration

Minimum	=	0
Maximum	=	154
Range	=	154
Sum	=	27855
Mean	=	76.3151
Median	=	78
Mode	=	66
Variance	=	823.1966
Standard deviation	=	28.6914
Standard error of the mean	=	1.5038
95 Percent confidence interval around the mean	=	73.3675 - 79.2626
Variance (unbiased)	=	825.4582
Standard deviation (unbiased)	=	28.7308
Skewness	=	0.0069
Kurtosis	=	2.5471
Kolmogorov-Smirnov statistic for normality	=	0.7927

Valid cases = 365
Missing cases = 0
Response percent = 100.0 %

StatPac Gold Statistical Analysis Package

DESCRIPTIVE STATISTICS FOR RUN 09

```
             33 I                *
             32 I
             31 I
             30 I
             29 I             *
             28 I
             27 I
             26 I
             25 I
             24 I
             23 I
             22 I                    *
             21 I              *
             20 I            *
             19 I                 *
 Number      18 I        * *       **
of Cases     17 I           *         *
             16 I         *
             15 I                      *
             14 I
             13 I
             12 I               *       *
             11 I       *                *
             10 I
              9 I      *
              8 I
              7 I
              6 I
              5 I                         ***
              4 I    *
              3 I*    *
              2 I
              1 I *                          ** *
                -+---------+---------+---------+-
                 0        50        100       150

                          concentration
```

StatPac Gold Statistical Analysis Package

DESCRIPTIVE STATISTICS FOR RUN 10

concentration

Minimum = 127.1027

Maximum = 951.273

Range = 824.1703

Sum = 171967.2661

Mean = 471.1432

Median = 470.7827

Mode = Multi-Modal

Variance = 22402.5133

Standard deviation = 149.6747

Standard error of the mean = 7.8451

95 Percent confidence interval around the mean = 455.7668 - 486.5196

Variance (unbiased) = 22464.0586

Standard deviation (unbiased) = 149.8801

Skewness = 0.2965

Kurtosis = 2.7661

Kolmogorov-Smirnov statistic for normality = 0.8333

Valid cases = 365
Missing cases = 0
Response percent = 100.0 %

StatPac Gold Statistical Analysis Package

DESCRIPTIVE STATISTICS FOR RUN 10

```
              37 I               *
              36 I
              35 I
              34 I
              33 I
              32 I
              31 I
              30 I            *
              29 I
              28 I
              27 I
              26 I                 *
              25 I
              24 I
              23 I
              22 I                *
              21 I         *
 Number       20 I        *     *
of Cases      19 I                  *
              18 I
              17 I      **
              16 I
              15 I             *        *
              14 I          **
              13 I
              12 I                     *
              11 I
              10 I
               9 I                       *
               8 I
               7 I   * *
               6 I    *              **
               5 I                         *
               4 I                        *  *
               3 I                          *   *
               2 I  *
               1 I**                           *    *
                 ------+---------+---------+--------
                      250       500       750

                             concentration
```

StatPac Gold Statistical Analysis Package

DESCRIPTIVE STATISTICS FOR RUN 11

concentration

```
     Minimum                    =  779.9222

     Maximum                    =  4563.007

     Range                      =  3783.0848

     Sum                        =  454213.9525

     Mean                       =  1244.4218

     Median                     =  1081.9090

     Mode                       =  Multi-Modal

     Variance                   =  239345.6321

     Standard deviation         =  489.2296

     Standard error of the mean  =  25.6426

     95 Percent confidence interval around the mean =  1194.1624 - 1294.6813

     Variance (unbiased)                =  240003.1750

     Standard deviation (unbiased)      =  489.9012

     Skewness      =  3.2702

     Kurtosis      =  15.4785

     Kolmogorov-Smirnov statistic for normality =  5.1572

Valid cases      = 365
Missing cases    = 0
Response percent = 100.0 %
```

StatPac Gold Statistical Analysis Package

DESCRIPTIVE STATISTICS FOR RUN 11

```
            110 I   *
            105 I    *
            100 I
             95 I
             90 I
             85 I
             80 I
             75 I
             70 I
             65 I
 Number      60 I
of Cases     55 I
             50 I
             45 I
             40 I
             35 I
             30 I     *
             25 I  *
             20 I      **
             15 I        *
             10 I
              5 I *       *   **          *
              0 I          **   ***** * *      **       *
                ----+---------+---------+---------+------
                  1000      2000      3000      4000

                              concentration
```

StatPac Gold Statistical Analysis Package

DESCRIPTIVE STATISTICS FOR RUN 12

concentration

Minimum	=	0
Maximum	=	1027.326
Range	=	1027.3260
Sum	=	82658.5856136
Mean	=	226.4619
Median	=	190.1387
Mode	=	0
Variance	=	55686.1177
Standard deviation	=	235.9791

Standard error of the mean = 12.3687

95 Percent confidence interval around the mean = 202.2193 - 250.7045

Variance (unbiased) = 55839.1015

Standard deviation (unbiased) = 236.3030

Skewness = 0.7969

Kurtosis = 2.7959

Kolmogorov-Smirnov statistic for normality = 3.8969

Valid cases = 365
Missing cases = 0
Response percent = 100.0 %

StatPac Gold Statistical Analysis Package

DESCRIPTIVE STATISTICS FOR RUN 12

```
            125 I*
            120 I
            115 I
            110 I
            105 I
            100 I
             95 I
             90 I
             85 I
             80 I
             75 I
             70 I
 Number      65 I
of Cases     60 I
             55 I
             50 I
             45 I
             40 I
             35 I
             30 I *
             25 I
             20 I        *
             15 I         *       *
             10 I      *   * ****  * ** *
              5 I    ** *   *    *  *  * ** ** *
              0 I  *                       *  *   * *   ***
                -+---------+---------+---------+---------+-
                 0        250       500       750      1000

                               concentration
```

StatPac Gold Statistical Analysis Package

DESCRIPTIVE STATISTICS FOR RUN 13

concentration

```
    Minimum                  =  2.522949

    Maximum                  =  4857.689

    Range                    =  4855.1661

    Sum                      =  1033004.893779

    Mean                     =  2830.1504

    Median                   =  2901.6680

    Mode                     =  Multi-Modal

    Variance                 =  919733.4721

    Standard deviation       =  959.0274

    Standard error of the mean  =  50.2667

    95 Percent confidence interval around the mean =  2731.6277 - 2928.6731

    Variance (unbiased)             =  922260.2124

    Standard deviation (unbiased)   =  960.3438

    Skewness      = -0.3201

    Kurtosis      =  2.7190

    Kolmogorov-Smirnov statistic for normality =  1.0032

Valid cases      = 365
Missing cases    = 0
Response percent = 100.0 %
```

StatPac Gold Statistical Analysis Package

DESCRIPTIVE STATISTICS FOR RUN 13

```
              23 I                     *
                 I
              22 I
                 I
              21 I
                 I
              20 I
                 I
              19 I                               *
                 I
              18 I
                 I
              17 I                         *    *       *
                 I
              16 I                                     *
                 I
              15 I
                 I
              14 I                          **     *     *
                 I
              13 I               *
 Number          I
of Cases      12 I                      *      *     *
                 I
              11 I                                  *
                 I
              10 I                *   *           *
                 I
               9 I                   *                       *
                 I
               8 I                  *                      *
                 I
               7 I                            *       *   *
                 I
               6 I             *          *                   *
                 I
               5 I
                 I
               4 I*                *     *
                 I
               3 I         * *  *                           *
                 I
               2 I       *                                     ***
                 I
               1 I *        * *                                   **
                 -+---------+---------+---------+---------+---------
                  0       1000      2000      3000      4000

                                    concentration
```

tatPac Gold Statistical Analysis Package

ESCRIPTIVE STATISTICS FOR RUN 14

oncentration

Minimum = 36

Maximum = 249.3818

Range = 213.3818

Sum = 74652.80760000001

Mean = 204.5282

Median = 206.0482

Mode = Multi-Modal

Variance = 275.0866

Standard deviation = 16.5857

Standard error of the mean = 0.8693

95 Percent confidence interval around the mean = 202.8244 - 206.2321

Variance (unbiased) = 275.8423

Standard deviation (unbiased) = 16.6085

Skewness = -3.0920

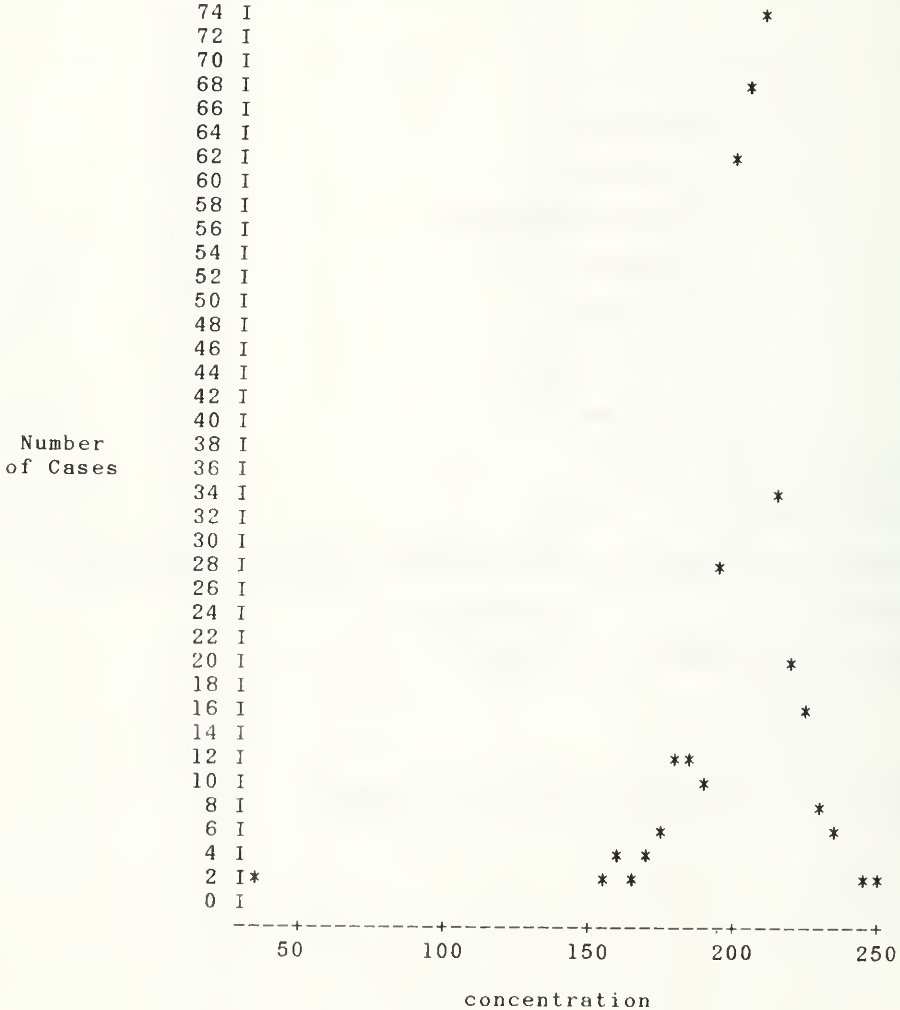
Kurtosis = 31.5423

Kolmogorov-Smirnov statistic for normality = 2.5905

alid cases = 365
issing cases = 0
esponse percent = 100.0 %

StatPac Gold Statistical Analysis Package

DESCRIPTIVE STATISTICS FOR RUN 14



StatPac Gold Statistical Analysis Package

DESCRIPTIVE STATISTICS FOR RUN 15

concentration

Minimum = 910.9802

Maximum = 12340

Range = 11429.0198

Sum = 583891.5258

Mean = 1599.7028

Median = 1546.9790

Mode = Multi-Modal

Variance = 395959.8395

Standard deviation = 629.2534

Standard error of the mean = 32.9818

95 Percent confidence interval around the mean = 1535.0583 - 1664.3472

Variance (unbiased) = 397047.6413

Standard deviation (unbiased) = 630.1172

Skewness = 13.6274

Kurtosis = 232.6578

Kolmogorov-Smirnov statistic for normality = 3.8268

Valid cases = 365
Missing cases = 0
Response percent = 100.0 %

StatPac Gold Statistical Analysis Package

DESCRIPTIVE STATISTICS FOR RUN 15

```
            130 I    *
            125 I
            120 I
            115 I
            110 I
            105 I
            100 I
             95 I
             90 I     *
             85 I
             80 I
             75 I   *
  Number     70 I
of Cases     65 I
             60 I
             55 I
             50 I
             45 I      *
             40 I
             35 I
             30 I
             25 I
             20 I
             15 I  *
             10 I
              5 I       *
              0 I        **                                    *
                --------+---------+---------+---------+---------
                      2500      5000      7500     10000

                                 concentration
```

StatPac Gold Statistical Analysis Package

DESCRIPTIVE STATISTICS FOR RUN 16

concentration

```
    Minimum                  =  0

    Maximum                  =  140

    Range                    =  140

    Sum                      =  25863

    Mean                     =  70.8575

    Median                   =  69

    Mode                     =  0

    Variance                 =  1075.2509

    Standard deviation       =  32.7910

    Standard error of the mean  =  1.7187

    95 Percent confidence interval around the mean =  67.4889 - 74.2262

    Variance (unbiased)              =  1078.2049

    Standard deviation (unbiased)    =  32.8360

    Skewness       = -0.2098

    Kurtosis       =  2.5946

    Kolmogorov-Smirnov statistic for normality =  0.8071

Valid cases      = 365
Missing cases    = 0
Response percent = 100.0 %
```

StatPac Gold Statistical Analysis Package

DESCRIPTIVE STATISTICS FOR RUN 16

```
              29 I             *  *
              28 I
              27 I
              26 I
              25 I
              24 I
              23 I
              22 I
              21 I*        * *
              20 I
              19 I            * *
              18 I          *      *
              17 I                    *
Number        16 I       *          *
of Cases      15 I                     *
              14 I                        *
              13 I        *          *
              12 I                      **
              11 I
              10 I
               9 I               *
               8 I
               7 I
               6 I                         *
               5 I    *                     **
               4 I     **
               3 I
               2 I  *
               1 I   *                        *
                 -+---------+---------+--------
                  0        50        100

                        concentration
```

StatPac Gold Statistical Analysis Package

DESCRIPTIVE STATISTICS FOR RUN 17

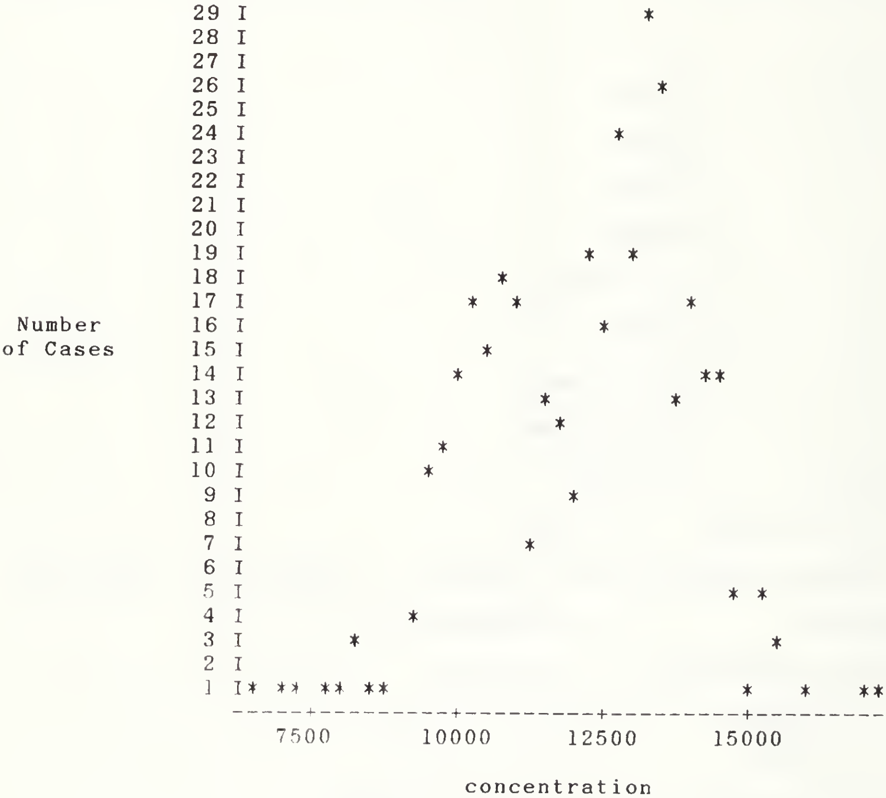
concentration

Minimum	=	6578
Maximum	=	17146
Range	=	10568
Sum	=	4449681
Mean	=	12190.9068
Median	=	12445
Mode	=	Multi-Modal
Variance	=	3040268.0845
Standard deviation	=	1743.6365
Standard error of the mean	=	91.3914
95 Percent confidence interval around the mean	=	12011.7793 - 12370.0342
Variance (unbiased)	=	3048620.4693
Standard deviation (unbiased)	=	1746.0299
Skewness	=	-0.2835
Kurtosis	=	2.7637
Kolmogorov-Smirnov statistic for normality	=	1.5542

Valid cases = 365
Missing cases = 0
Response percent = 100.0 %

StatPac Gold Statistical Analysis Package

DESCRIPTIVE STATISTICS FOR RUN 17



StatPac Gold Statistical Analysis Package

DESCRIPTIVE STATISTICS FOR RUN 18

concentration

Minimum	=	7.276834
Maximum	=	102.7069
Range	=	95.4301
Sum	=	17458.407344
Mean	=	47.8313
Median	=	47.0727
Mode	=	Multi-Modal
Variance	=	150.4505
Standard deviation	=	12.2658

Standard error of the mean = 0.6429

95 Percent confidence interval around the mean = 46.5712 - 49.0913

Variance (unbiased) = 150.8638

Standard deviation (unbiased) = 12.2827

Skewness = 0.3936

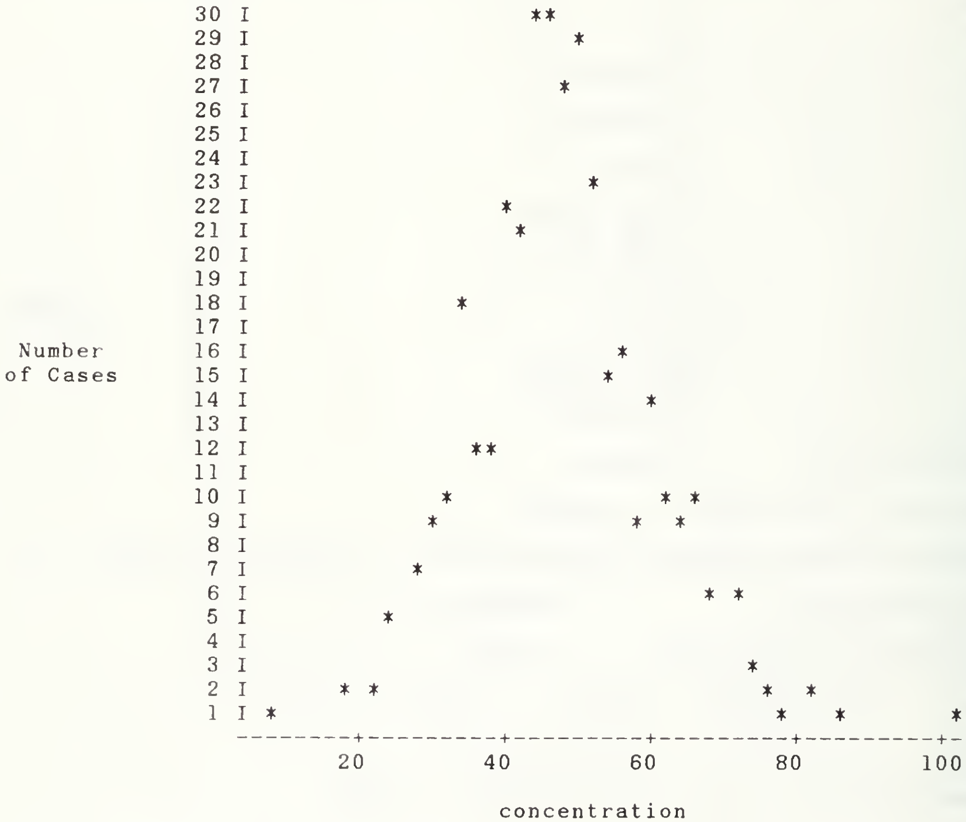
Kurtosis = 4.0401

Kolmogorov-Smirnov statistic for normality = 1.1431

Valid cases = 365
Missing cases = 0
Response percent = 100.0 %

StatPac Gold Statistical Analysis Package

DESCRIPTIVE STATISTICS FOR RUN 18



StatPac Gold Statistical Analysis Package

DESCRIPTIVE STATISTICS FOR RUN 19

concentration

Minimum = 67.08031

Maximum = 1300

Range = 1232.9197

Sum = 47272.92964

Mean = 129.5149

Median = 128.9166

Mode = Multi-Modal

Variance = 4113.6958

Standard deviation = 64.1381

Standard error of the mean = 3.3617

95 Percent confidence interval around the mean = 122.9258 - 136.1039

Variance (unbiased) = 4124.9972

Standard deviation (unbiased) = 64.2261

Skewness = 16.6282

Kurtosis = 303.9075

Kolmogorov-Smirnov statistic for normality = 6.3586

Valid cases = 365
Missing cases = 0
Response percent = 100.0 %

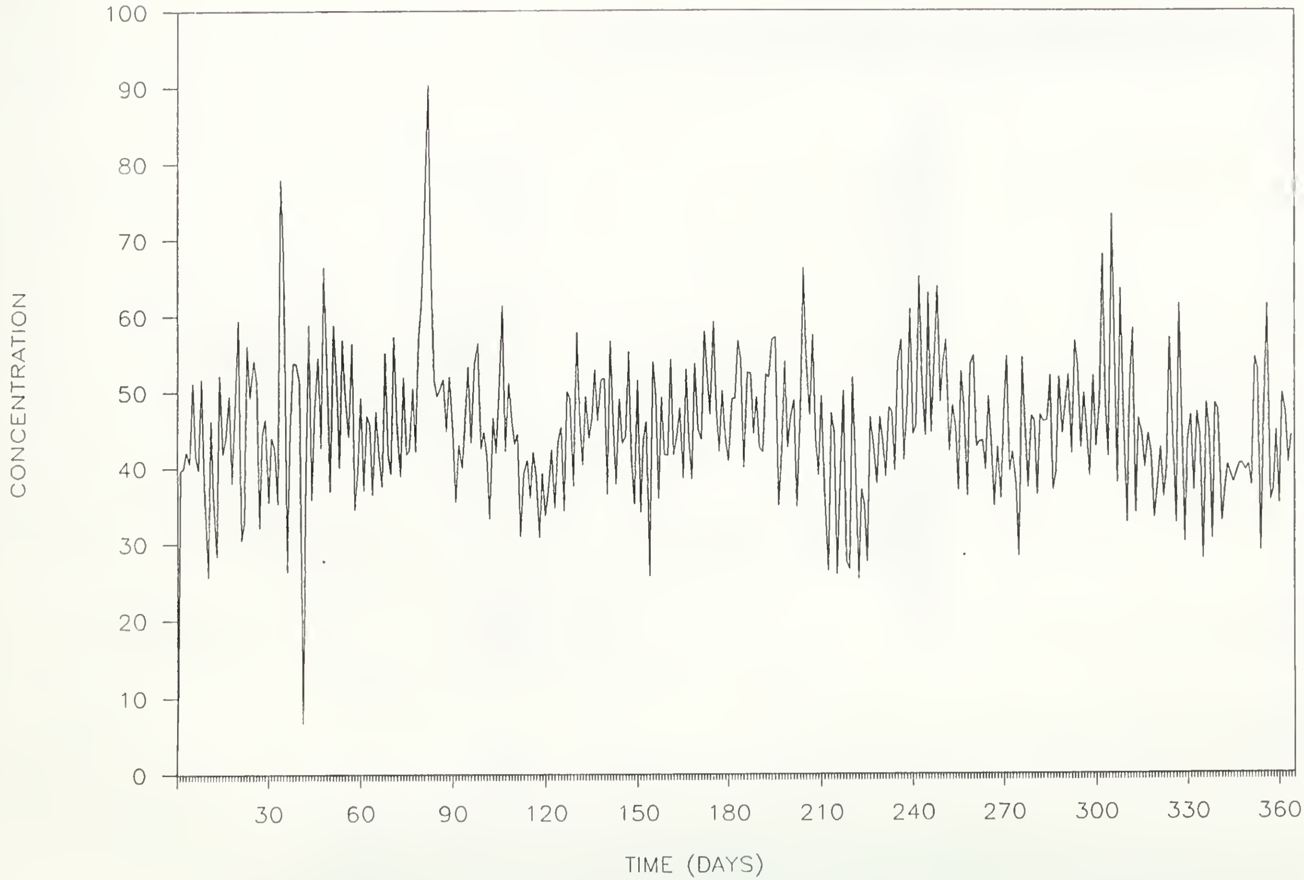
StatPac Gold Statistical Analysis Package

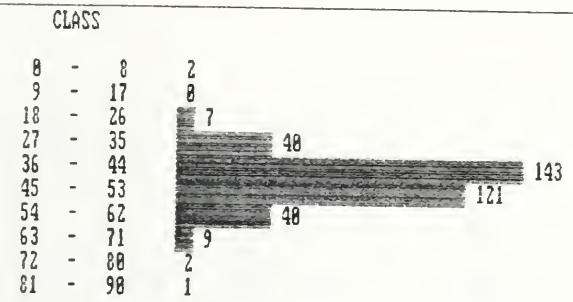
DESCRIPTIVE STATISTICS FOR RUN 19

```
            145 I   *
            140 I
            135 I
            130 I
            125 I
            120 I    *
            115 I
            110 I
            105 I
            100 I
             95 I
             90 I
             85 I
             80 I  *
 Number      75 I
of Cases     70 I
             65 I
             60 I
             55 I
             50 I
             45 I
             40 I
             35 I
             30 I
             25 I
             20 I
             15 I
             10 I *
              5 I     *
              0 I                                                  *
                ---------+---------+---------+---------+---------+--
                        250       500       750      1000      1250

                                   concentration
```

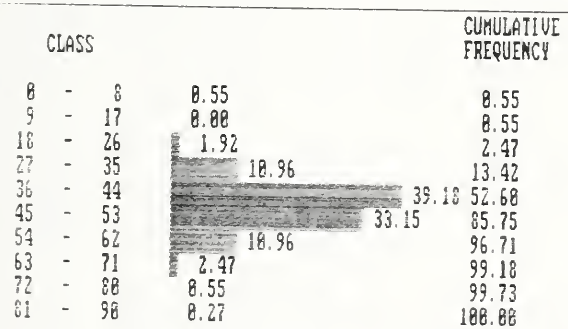
C3: DATA USED TO REPRESENT DIFFERENT INDUSTRIAL VARIABILITY LEVELS

TIME SERIES SCATTERPLOT FOR RUN VL





A B S O L U T E F R E Q U E N C Y D I S T R I B U T I O N
LOW VARIABILITY INDUSTRY (DATA00VL.PRN)



R E L A T I V E F R E Q U E N C Y D I S T R I B U T I O N
LOW VARIABILITY INDUSTRY (DATA00VL.PRN)

[illegible]

[illegible] DATA[illegible].PR

THE MEAN IS 45.190
THE SD I[illegible] [illegible]
THE MINIMUM IS 0.000
THE MAXIMUM IS [illegible]
RANGE IS [illegible]
THE VARIANCE OF X IS [illegible]
THE SKEWNESS COEFFICIE[illegible] [illegible]
THE KURTOSIS COEFFICIENT IS [illegible]
THE EXCESS COEFFICIENT I[illegible] [illegible]
THE COEFFICIENT OF VARIATION [illegible]
[illegible]

[illegible]

[illegible] DATA[illegible]

[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	44	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	54	56	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	42	[illegible]	61	[illegible]	51
[illegible]	44	31	[illegible]	[illegible]	[illegible]	42	39	31	39	34	[illegible]	42	35	43	45	34	50	49	38	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
26	54	49	[illegible]	49	42	42	54	42	44	48	39	53	44	38	54	45	44	58	52	47
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	42	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
54	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	54	47	57	44	39	49	[illegible]	[illegible]	47	45	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	43	39	46	[illegible]	40	54	57	41	[illegible]	[illegible]	44
[illegible]	[illegible]	[illegible]	[illegible]	45	[illegible]	[illegible]	49	54	57	[illegible]	[illegible]	45	[illegible]	53	48	[illegible]	54	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	55	35	[illegible]	38	28	54	46	37	[illegible]	46	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	57	[illegible]	[illegible]	50	45	[illegible]	52	43	49	68	[illegible]	41	[illegible]	[illegible]
64	[illegible]	[illegible]	52	58	[illegible]	[illegible]	44	[illegible]	[illegible]	42	[illegible]	[illegible]	43	[illegible]	40	57	45	[illegible]	62	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	40
55	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	45	[illegible]	[illegible]	[illegible]	41	44								

* * * *CONFIDENCE INTERVAL ANALYSIS * * * *
GARTNER LEE LIMITED
SIMULATION RUN NUMBER VL
11-21-1988
21:25:46
STATISTICS CALCULATED USING POPULATION MEAN (Mu)

DATASET c:\DESIGN\DATA00VL.PRN

MONTH 1

FR	N	UCL	MEAN	LCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
D	30	45.76	41.54	37.31	8.44	11.34	128.52	-0.05	-1.49	3.71	0.00	1
T	13	48.20	43.31	38.41	9.79	8.10	65.56	0.07	-0.13	-1.05	4.26	2
W	4	59.58	40.41	21.24	38.34	12.06	145.34	-0.71	0.38	-1.34	-2.71	7

MONTH 2

FR	N	UCL	MEAN	LCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
D	30	51.85	46.86	41.88	9.97	13.39	179.27	-0.04	-0.40	1.39	0.00	1
T	13	52.01	46.29	40.58	11.43	9.45	89.39	0.26	-0.68	-0.59	-1.22	2
W	4	93.45	50.08	6.71	86.75	27.28	744.13	-0.66	-0.34	-1.33	6.86	7

MONTH 3

FR	N	UCL	MEAN	LCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
D	30	54.04	49.85	45.65	8.40	11.27	127.06	0.54	1.69	3.52	0.00	1
T	13	58.80	50.40	42.00	16.80	13.89	192.97	0.27	1.75	2.77	1.12	2
W	4	66.20	49.64	32.99	33.29	10.47	109.61	-0.23	0.98	-0.85	-0.42	7

MONTH 4

FR	N	UCL	MEAN	LCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
D	30	46.05	43.47	40.89	5.16	6.93	48.03	0.27	0.46	0.27	0.00	1
T	12	48.52	44.44	40.37	8.15	6.42	41.19	0.49	1.79	2.43	2.24	2
W	5	52.29	43.16	34.03	18.26	7.34	53.94	-0.13	-0.47	-0.50	-0.71	7

MONTH 5

FR	N	UCL	MEAN	LCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
D	30	47.17	44.64	42.12	5.05	6.77	45.90	-0.03	0.11	-0.91	0.00	1
T	13	49.08	44.42	39.75	9.33	7.72	59.59	0.34	-0.01	-1.24	-0.51	2
W	4	52.03	45.38	38.74	13.28	4.18	17.45	-0.58	0.95	-0.63	1.66	7

MONTH 6

FR	N	UCL	MEAN	LCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
D	30	48.55	45.89	43.23	5.32	7.14	50.94	-0.25	-0.49	0.49	0.00	1
T	13	50.29	44.75	39.21	11.07	9.16	83.86	0.35	-0.60	-0.29	-2.48	2
W	4	48.51	44.55	40.59	7.92	2.49	6.21	-0.01	-1.45	-0.63	-2.92	7

MONTH 7

FR	N	UCL	MEAN	LCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
D	30	50.99	48.32	45.64	5.35	7.18	51.52	0.09	0.15	-0.28	0.00	1
T	12	53.21	49.41	45.60	7.61	5.99	35.88	-0.21	-0.43	-0.25	2.26	2
W	5	55.51	44.90	34.28	21.23	8.54	72.89	-0.29	-0.84	-1.22	-7.07	7

MONTH 8

FR	N	UCL	MEAN	LCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
D	30	44.88	41.35	37.82	7.06	9.48	89.79	0.16	-0.13	-0.73	0.00	1
T	13	48.19	41.19	34.19	14.00	11.58	134.07	0.10	-0.02	-1.24	-0.39	2
W	4	44.46	39.88	35.30	9.16	2.88	8.31	-0.11	-1.08	-1.33	-3.55	7

MONTH 9

FR	N	UCL	MEAN	LCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
D	30	50.53	47.58	44.64	5.89	7.91	62.54	0.11	0.53	-0.31	0.00	1
T	13	51.11	46.16	41.22	9.89	8.18	66.91	0.33	-0.08	-0.68	-2.99	2
W	4	[illegible]	46.63	34.62	24.01	7.55	57.02	-0.73	-0.41	-1.16	-2.01	7

MONTH 10

FR	N	UCL	MEAN	LCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
D	30	47.75	45.30	42.86	4.88	6.55	42.92	-0.09	-0.37	-0.26	0.00	1
T	13	50.02	46.13	42.25	7.76	6.42	41.21	-0.05	0.21	-1.44	1.83	2
W	4	[illegible]	[illegible]	[illegible]	24.38	7.67	58.76	-0.77	0.38	-1.36	-0.84	7

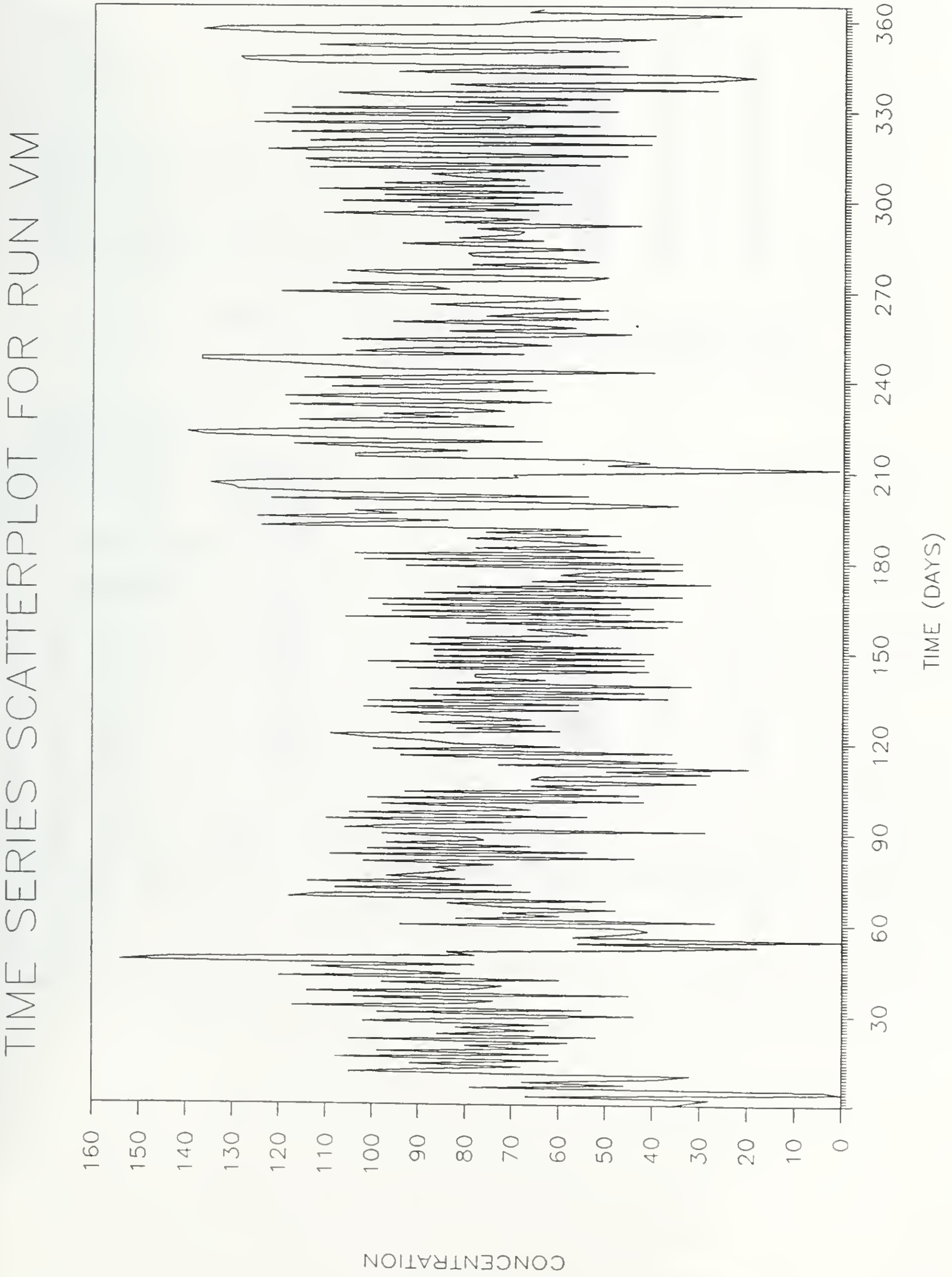
MONTH 11

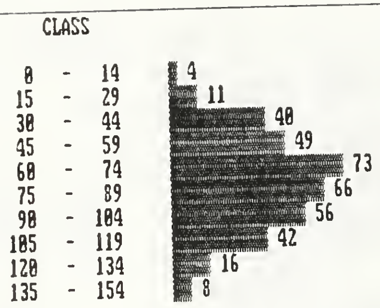
FR	N	UCL	MEAN	LCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
D	30	50.00	46.01	42.01	7.99	10.72	114.93	-0.03	0.78	-0.03	0.00	1
T	12	53.05	46.82	40.60	12.45	9.81	96.15	0.26	1.09	-0.07	1.78	2
W	5	48.46	43.19	37.92	10.53	4.24	17.95	-0.29	-1.62	-0.11	-6.12	7

MONTH 12

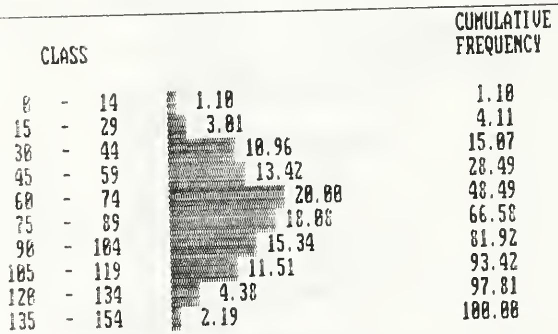
FR	N	UCL	MEAN	LCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
D	30	44.52	41.77	39.03	5.49	7.37	54.26	-0.17	0.39	0.26	0.00	1
T	13	46.08	41.80	37.51	8.58	7.09	50.31	-0.27	-0.20	-0.70	0.05	2
W	4	61.06	41.46	21.86	39.19	12.32	151.90	-0.39	0.64	-0.97	-0.75	7

Computation time = 3.0 minutes.


TIME SERIES SCATTERPLOT FOR RUN VM
CONCENTRATION
TIME (DAYS)
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160
30 60 90 120 150 180 210 240 270 300 330 360



A B S O L U T E   F R E Q U E N C Y   D I S T R I B U T I O N
MEDIUM VARIABILITY INDUSTRY (DATA00VM.PRN)



R E L A T I V E   F R E Q U E N C Y   D I S T R I B U T I O N
MEDIUM VARIABILITY INDUSTRY (DATA00VM.PRN)

09-14-1988 14:12:04

D E S C R I P T I V E   S T A T I S T I C S

C:\design\DATA00VM.PRN

THE MEAN IS	76.315
THE SD IS	28.691
THE MINIMUM IS	0.000
THE MAXIMUM IS	154.000
RANGE IS	154.000
THE VARIANCE OF X IS	823.197
THE SKEWNESS COEFFICIENT IS	0.0069
THE KURTOSIS COEFFICIENT IS	2.5471
THE EXCESS COEFFICIENT IS	-0.4529
THE COEFFICIENT OF VARIATION IS	37.6
AUTOCORR. COEFFICENT (LAG=1) IS	0.0755

* * * *CONFIDENCE INTERVAL ANALYSIS * * * *
GARTNER LEE LIMITED
SIMULATION RUN NUMBER VM
09-14-1988
14:27:45
STATISTICS CALCULATED USING POPULATION MEAN (MU)

DATASET c:\DESIGN\DATA00VM.PRN

36	32	25	67	0	4	79	46	68	39	32	105	90	68	92	60	105	62	99	66	80	[illegible]
105	52	86	65	92	62	102	83	44	94	55	117	80	74	104	45	114	78	72	98	60	[illegible]
81	96	113	[illegible]	[illegible]	143	78	84	40	18	56	0	57	46	41	44	44	27	82	60	72	[illegible]
72	84	50	[illegible]	112	66	[illegible]	70	114	80	97	98	82	67	74	102	44	109	54	101	66	[illegible]
76	78	96	29	106	94	70	110	54	105	66	74	78	42	101	43	93	52	66	31	66	64
28	50	20	77	35	66	94	36	100	60	83	88	102	109	60	82	65	90	66	77	96	54
102	56	101	37	87	42	92	32	82	63	76	76	41	95	42	101	42	67	40	67	47	[illegible]
62	88	54	60	57	37	80	34	106	45	96	40	93	47	101	34	89	48	82	28	66	[illegible]
60	54	34	43	34	102	40	104	43	76	50	62	80	47	76	54	124	107	64	125	95	104
54	35	60	122	54	106	129	130	135	122	69	70	33	1	50	41	43	104	104	80	96	117
64	108	136	140	114	70	85	116	82	96	72	81	115	62	88	119	88	63	109	83	66	115
82	40	99	109	121	137	137	68	104	94	62	74	107	63	45	84	57	66	96	50	76	65
50	74	88	[illegible]	[illegible]	78	120	34	88	104	54	50	95	106	54	79	52	60	79	80	55	75
94	64	82	69	68	78	43	85	67	86	111	65	91	58	107	66	98	60	112	67	98	[illegible]
80	89	64	114	52	107	115	46	85	120	103	41	114	92	40	118	92	52	126	72	71	124
48	118	59	97	50	96	108	27	71	84	31	19	27	95	78	46	116	128	129	80	48	[illegible]
112	57	40	69	121	137	132	74	68	40	22	67	64									

SIMULATION RUN NUMBER VM
09-14-1988
14:28:23

MONTH 1

FR	N	UCL	MEAN	LCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
D	30	75.75	65.20	54.65	21.09	28.32	801.89	0.15	-0.49	-0.37	0.00	1
T	13	80.50	64.38	48.27	32.22	26.65	709.98	-0.05	-0.83	0.01	-1.25	2
W	4	88.69	72.25	55.81	32.88	10.34	106.89	-0.67	1.33	-1.15	10.81	7

MONTH 2

FR	N	UCL	MEAN	LCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
D	30	89.27	76.30	63.33	25.93	34.81	1212.08	0.31	0.14	-0.25	0.00	1
T	13	98.93	81.85	64.77	34.16	28.25	797.97	0.47	0.43	-1.40	7.27	2
W	4	171.33	85.00	0.00	176.65	55.55	3085.89	-0.38	0.05	-1.15	8.78	7

MONTH 3

FR	N	UCL	MEAN	LCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
D	30	88.69	80.40	72.11	16.58	22.26	495.37	-0.37	-0.30	-0.46	0.00	1
T	13	83.55	70.23	56.91	26.63	22.03	485.13	-0.01	-1.59	0.62	-12.65	2
W	4	141.33	101.50	61.67	79.66	25.05	627.46	-0.43	1.29	-1.26	26.24	7

MONTH 4

FR	N	UCL	MEAN	LCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
D	30	77.46	67.47	57.47	19.99	26.83	720.12	-0.32	-0.02	-1.23	0.00	1
T	12	73.99	60.33	46.68	27.31	21.50	462.27	-0.28	-0.57	-0.49	-10.57	2
W	5	125.90	89.20	52.50	73.40	29.52	871.30	-0.09	0.99	-1.82	32.21	7

MONTH 5

FR	N	UCL	MEAN	LCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
D	30	82.66	74.33	66.01	16.65	22.35	499.36	-0.57	-0.38	-1.09	0.00	1
T	13	86.68	72.77	58.85	27.83	23.02	529.70	0.11	-0.35	-1.18	-2.10	2
W	4	112.61	64.50	16.39	96.22	30.26	915.44	-0.06	-0.79	-1.54	-13.23	7

MONTH 6

FR	N	UCL	MEAN	LCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
D	30	72.54	63.63	54.73	17.81	23.90	571.37	-0.74	0.25	-1.36	0.00	1
T	13	78.57	65.23	51.89	26.68	22.06	486.73	0.24	0.37	-1.36	2.51	2
W	4	100.51	71.25	41.99	58.52	18.40	338.70	-0.01	0.91	-1.27	11.97	7

MONTH 7

FR	N	UCL	MEAN	LCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
D	30	95.04	83.17	71.29	23.75	31.88	1016.27	0.20	0.09	-1.36	0.00	1
T	12	96.61	75.67	54.72	41.89	32.98	1087.81	0.41	-0.24	-1.70	-9.02	2
W	5	117.35	86.20	55.05	62.30	25.06	627.76	-0.42	0.67	-1.00	3.65	7

MONTH 8

FR	N	UCL	MEAN	LCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
D	30	97.34	85.77	74.19	23.15	31.08	966.05	0.44	-0.56	0.15	0.00	1
T	13	105.56	85.00	64.44	41.13	34.01	1156.74	0.11	-1.01	0.70	-0.89	2
W	4	144.91	94.75	44.59	100.33	31.55	995.39	-0.05	1.20	-0.71	10.47	7

MONTH 9

FR	N	UCL	MEAN	LCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
D	30	90.54	81.00	71.46	19.09	25.63	656.67	0.28	0.56	-0.48	0.00	1
T	13	101.74	85.08	68.41	33.32	27.56	759.46	0.07	0.69	-0.61	5.03	2
W	4	107.58	81.00	54.42	53.16	16.72	279.50	-0.03	-0.33	-1.50	0.00	7

MONTH 10

FR	N	UCL	MEAN	LCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
D	30	84.03	76.87	69.70	14.33	19.23	369.85	-0.15	0.36	-0.57	0.00	1
T	13	79.63	69.92	60.21	19.42	16.06	257.82	-0.45	-1.17	-0.64	-9.03	2
W	4	104.07	86.50	68.93	35.13	11.05	122.05	-0.34	1.29	-1.22	12.53	7

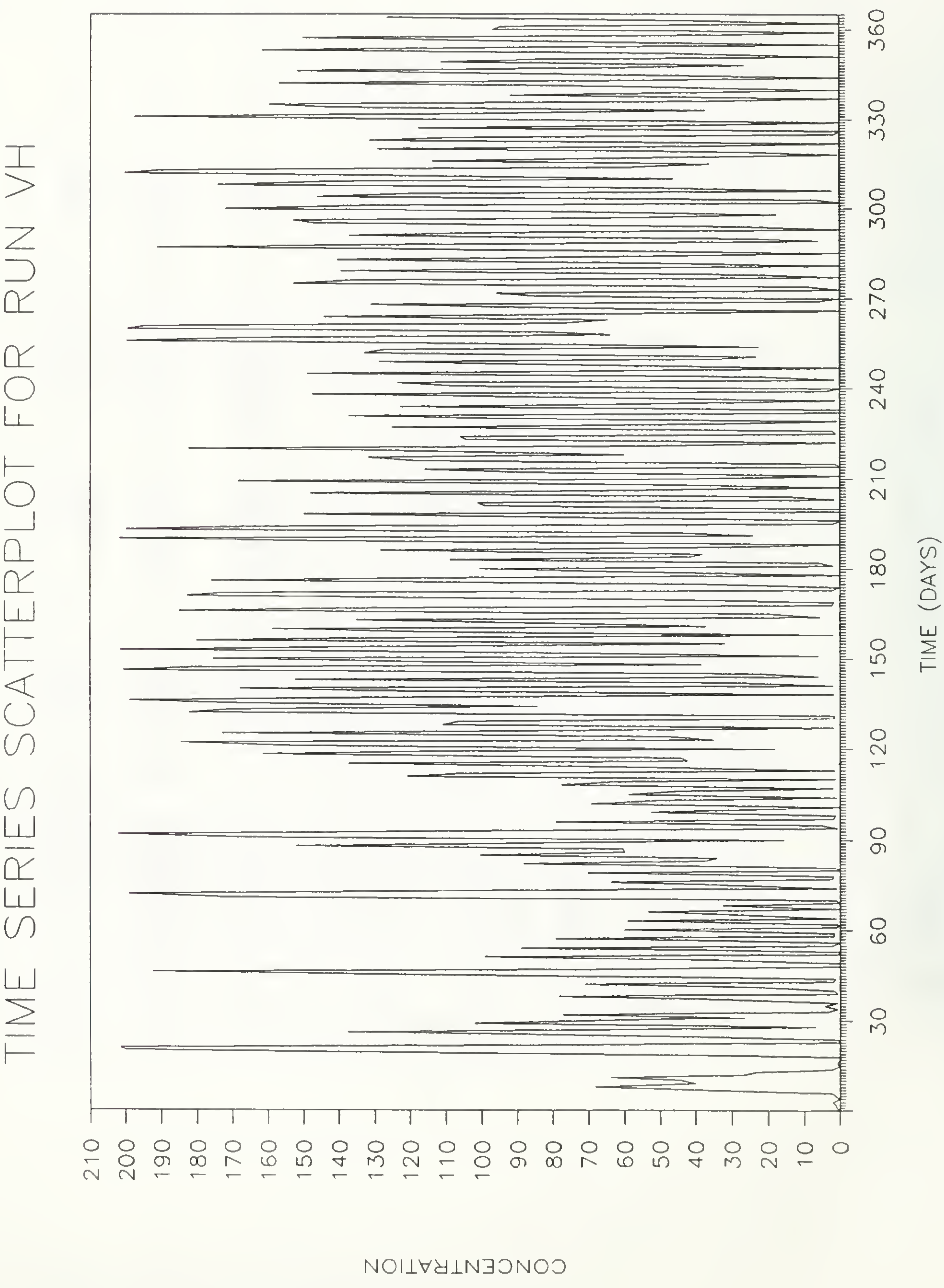
MONTH 11

FR	N	UCL	MEAN	LCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
D	30	96.35	86.53	76.72	19.64	26.36	694.92	-0.47	-0.19	-1.23	0.00	1
T	12	110.49	96.83	83.17	27.32	21.51	462.56	0.10	0.52	-1.31	11.90	2
W	5	114.47	90.60	66.73	47.75	19.20	368.78	-0.69	0.61	-1.52	4.70	7

MONTH 12

FR	N	UCL	MEAN	LCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
D	30	92.06	79.13	66.20	25.86	34.71	1204.92	0.28	0.02	-1.16	0.00	1
T	13	104.76	83.92	63.09	41.67	34.46	1187.32	-0.61	0.36	-1.31	6.05	2
W	4	135.83	81.00	26.17	109.65	34.48	1188.98	-0.04	-0.39	-1.14	2.36	7

Computation time = 3.4 minutes.


TIME SERIES SCATTERPLOT FOR RUN VH
CONCENTRATION
TIME (DAYS)
0
10
20
30
40
50
60
70
80
90
100
110
120
130
140
150
160
170
180
190
200
210
30
60
90
120
150
180
210
240
270
300
330
360

CLASS	
0 - 19	118
20 - 39	36
40 - 59	35
60 - 79	37
80 - 99	22
100 - 119	30
120 - 139	26
140 - 159	21
160 - 179	16
180 - 202	24

A B S O L U T E   F R E Q U E N C Y   D I S T R I B U T I O N
HIGH VARIABILITY INDUSTRY (DATA00VH.PRN)

CLASS		CUMULATIVE FREQUENCY
0 - 19	32.33	32.33
20 - 39	9.86	42.19
40 - 59	9.59	51.78
60 - 79	10.14	61.92
80 - 99	6.03	67.95
100 - 119	8.22	76.16
120 - 139	7.12	83.29
140 - 159	5.75	89.04
160 - 179	4.38	93.42
180 - 202	6.58	100.00

R E L A T I V E   F R E Q U E N C Y   D I S T R I B U T I O N
HIGH VARIABILITY INDUSTRY (DATA00VH.PRN)

09-14-19[illegible] 14:44:5[illegible]

D E S C R I P T I V E   S T A T I S T I C S

C:\design\DATA00VB.PRN

THE MEAN IS	68.061
THE SD IS	62.622
THE MINIMUM IS	0.000
THE MAXIMUM IS	201.933
RANGE IS	201.933
THE VARIANCE OF X IS	3921.552
THE SKEWNESS COEFFICIENT IS	0.5435
THE KURTOSIS COEFFICIENT IS	2.0736
THE EXCESS COEFFICIENT IS	[illegible]
THE COEFFICIENT OF VARIATION	[illegible]
AUTOCORR. COEFFICENT [illegible]	[illegible]

* * * (CONFIDENCE INTERVAL A[illegible] * * * * *

GARDNER [illegible]

SIMULATION RUN NUMBER [illegible]

09 14-19[illegible]

14:45:[illegible]

STATISTICS CALCULATED [illegible] MEAN (M[illegible])

DATASET [illegible]

[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	7	1	3	1	37	78	1	1	23	71
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	59	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	34	100	60
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	45	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	6	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	185	[illegible]	2	[illegible]	62	183	173	2	0
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	116	[illegible]	[illegible]	[illegible]	132	60
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	4	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
114	[illegible]	[illegible]	[illegible]	[illegible]	6	0	35	9	[illegible]	[illegible]	157	135	[illegible]	25	179	85	3	[illegible]	140	56
76	191	115	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	27	147	153	56	18	85	172	131	0	7	146	119	2
174	143	46	11	[illegible]	190	59	36	114	57	0	27	129	42	0	131	110	1	0	118	42
101	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0	72	152	70	27	111	[illegible]
49	162	94	[illegible]	[illegible]	150	[illegible]	1	97	95	0	44	127								

* * * *CONFIDENCE INTERVAL ANALYSIS * * * *
GARTNER LEE LIMITED
SIMULATION RUN NUMBER VH
11-21-1988
21:29:16
STATISTICS CALCULATED USING POPULATION MEAN (MU)

DATASET c:\DESIGN\DATA00VH.PRN

MONTH 1

FR	N	UCL	MEAN	LCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
D	30	62.55	41.72	20.88	41.68	55.95	3130.34	0.55	1.58	1.85	0.00	1
T	13	69.33	41.83	14.33	55.00	45.48	2068.76	0.08	0.68	-0.84	0.27	2
W	4	201.77	70.76	0.00	262.02	82.40	6789.20	-0.14	1.74	0.41	69.62	7

MONTH 2

FR	N	UCL	MEAN	LCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
D	30	52.19	[illegible]	17.94	34.25	45.98	2113.95	0.11	1.49	2.19	0.00	1
T	13	59.90	36.91	13.93	45.97	38.02	1445.44	-0.11	0.62	-1.28	5.28	2
W	4	54.76	[illegible]	0.00	95.97	30.18	910.74	-0.13	-1.10	-1.77	-81.80	7

MONTH 3

FR	N	UCL	MEAN	LCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
D	30	69.9[illegible]	50.0[illegible]	30.23	39.71	53.31	2842.29	0.35	1.16	0.74	0.00	1
T	13	87.[illegible]	56.41	25.51	61.79	51.10	2611.29	0.23	1.00	-0.02	12.61	2
W	[illegible]	[illegible]	[illegible]	0.00	115.95	36.46	1329.08	-0.05	-1.27	-1.25	-49.32	7

MONTH 4

FR	N	UCL	MEAN	LCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
D	30	81.53	60.61	39.68	41.85	56.18	3156.34	0.19	0.78	-0.26	0.00	1
T	12	94.9[illegible]	[illegible]	20.27	74.68	58.80	3456.94	-0.35	0.99	0.38	-4.95	2
W	5	144.10	[illegible]	0.00	168.03	67.58	4566.65	0.18	0.48	-1.61	-0.87	7

MONTH 5

FR	N	UCL	MEAN	LCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
D	30	123.15	96.50	69.84	53.31	71.56	5121.15	-0.03	-0.06	-1.58	0.00	1
T	13	142.31	103.55	64.78	77.53	64.12	4111.12	-0.08	-0.07	-1.53	7.30	2
W	4	298.70	172.16	45.62	253.08	79.58	6333.75	-0.19	1.12	-1.71	78.41	7

MONTH 6

FR	N	UCL	MEAN	LCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
D	30	112.37	86.24	60.12	52.26	70.15	4921.49	-0.02	0.15	-1.50	0.00	1
T	13	112.03	73.82	35.61	76.42	63.20	3993.95	-0.18	-0.17	-1.48	-14.40	2
W	4	253.52	121.68	0.00	263.67	82.92	6875.06	-0.29	0.59	-1.62	41.09	7

MONTH 7

FR	N	UCL	MEAN	LCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
D	30	94.97	70.41	45.86	49.11	65.92	4345.73	0.03	0.52	-1.07	0.00	1
T	12	116.83	70.24	23.66	93.17	73.35	5380.32	-0.41	0.41	-1.52	-0.24	2
W	5	142.40	54.65	0.00	175.49	70.58	4981.06	0.11	-0.01	-1.70	-22.39	7

MONTH 8

FR	N	UCL	MEAN	LCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
D	30	84.90	63.86	42.83	42.07	56.48	3189.56	-0.04	0.23	-1.28	0.00	1
T	13	86.83	47.35	7.86	78.97	65.31	4265.30	-0.32	0.07	-1.50	-25.86	2
W	4	151.27	94.26	37.25	114.02	35.85	1285.53	-0.10	1.31	-1.20	47.60	7

MONTH 9

FR	N	UCL	MEAN	LCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
D	30	108.77	85.87	62.98	45.79	61.48	3779.31	0.15	0.24	-0.97	0.00	1
T	13	123.73	89.40	55.06	68.67	56.78	3224.45	0.11	0.42	-0.62	4.10	2
W	4	119.25	45.57	0.00	147.37	46.34	2147.58	-0.01	-1.29	-1.28	-46.93	7

MONTH 10

FR	N	UCL	MEAN	LCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
D	30	92.23	70.37	48.51	43.72	58.70	3445.44	0.04	0.28	-1.24	0.00	1
T	13	102.37	70.54	38.71	63.66	52.65	2771.68	-0.56	0.03	-1.47	0.24	2
W	4	159.37	78.38	0.00	161.97	50.93	2594.23	-0.03	0.00	-1.25	11.38	7

MONTH 11

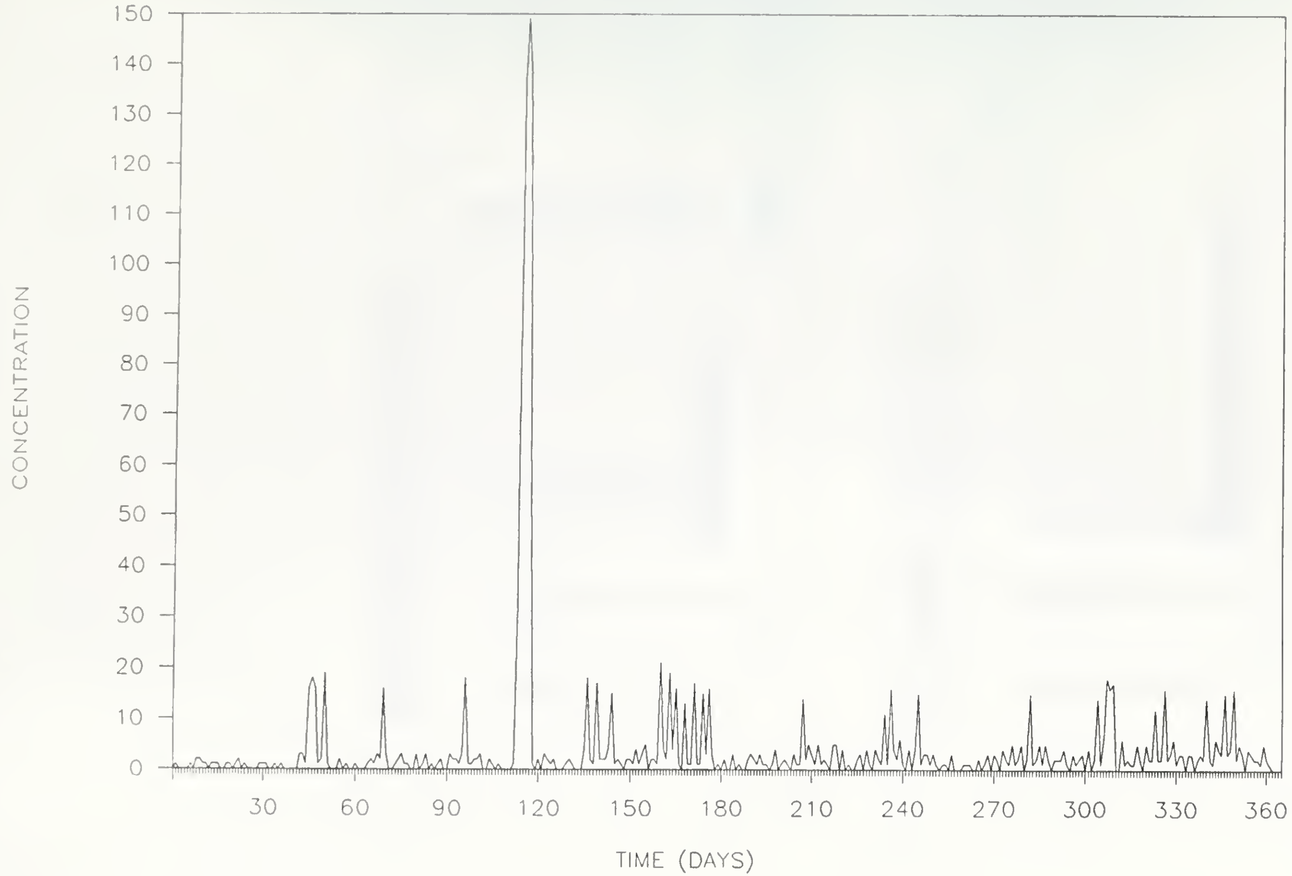
FR	N	UCL	MEAN	LCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
D	30	103.15	78.98	54.82	48.32	64.87	4208.58	0.12	0.23	-1.30	0.00	1
T	12	134.12	91.17	48.23	85.89	67.62	4572.63	-0.38	0.13	-1.57	15.43	2
W	5	135.62	73.71	11.81	123.81	49.79	2479.24	-0.27	1.10	-0.44	-6.67	7

MONTH 12

FR	N	UCL	MEAN	LCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
D	30	97.61	76.27	54.92	42.69	57.31	3284.95	0.05	0.31	-0.96	0.00	1
T	13	104.40	72.45	40.51	63.90	52.84	2792.06	-0.49	0.62	0.13	-5.00	2
W	4	225.61	128.83	32.05	193.55	60.87	3704.67	-0.02	1.27	-1.32	68.93	7

Computation time = 2.9 minutes.

TIME SERIES SCATTERPLOT FOR RUN VV



CLASS			
0	-	14	336
15	-	29	24
30	-	44	1
45	-	59	0
60	-	74	0
75	-	89	1
90	-	104	0
105	-	119	0
120	-	134	0
135	-	149	3

A B S O L U T E F R E Q U E N C Y D I S T R I B U T I O N
VERY HIGH VARIABILITY INDUSTRY (DATA00VV.PRN)

CLASS				CUMULATIVE FREQUENCY
0	-	14	92.05	92.05
15	-	29	6.58	98.63
30	-	44	0.27	98.90
45	-	59	0.00	98.90
60	-	74	0.00	98.90
75	-	89	0.27	99.18
90	-	104	0.00	99.18
105	-	119	0.00	99.18
120	-	134	0.00	99.18
135	-	149	0.82	100.00

R E L A T I V E F R E Q U E N C Y D I S T R I B U T I O N
VERY HIGH VARIABILITY INDUSTRY (DATA00VV.PRN)

11-21-1988 21:40:38

D E S C R I P T I V E   S T A T I S T I C S

C:\design\DATA00vv.PRN

THE MEAN IS	4.140
THE SD IS	13.852
THE MINIMUM IS	0.000
THE MAXIMUM IS	149.000
RANGE IS	149.000
THE VARIANCE OF X IS	191.868
THE SKEWNESS COEFFICIENT IS	8.4922
THE KURTOSIS COEFFICIENT IS	82.0592
THE EXCESS COEFFICIENT IS	79.0592
THE COEFFICIENT OF VARIATION IS	334.6
AUTOCORR. COEFFICENT (LAG=1) IS	0.7643

* * * *CONFIDENCE INTERVAL ANALYSIS * * * *
GARTNER LEE LIMITED
SIMULATION RUN NUMBER VV
11-21-1988
21:36:03
STATISTICS CALCULATED USING POPULATION MEAN (Mu)

DATASET c:\DESIGN\DATA00VV.PRN

0	1	0	0	0	0	1	0	2	2	1	1	0	1	1	1	0	0	1	1	0
2	0	1	0	0	0	0	1	1	1	0	0	1	0	1	0	0	0	0	0	3
1	16	18	16	1	2	19	1	0	0	0	2	0	1	0	0	1	0	0	0	1
1	3	2	16	3	0	0	1	2	3	1	1	0	0	3	0	1	3	0	1	0
2	0	0	3	2	2	1	3	18	1	1	2	2	3	0	0	2	1	0	1	0
0	0	1	33	137	149	138	77	1	0	2	0	3	2	1	2	0	0	0	1	2
0	0	0	4	18	2	1	17	2	2	2	4	15	1	2	1	0	2	2	1	4
3	5	0	2	2	1	21	4	2	19	4	16	1	0	13	1	1	17	1	1	15
16	3	0	1	0	2	0	0	3	0	1	0	0	2	3	2	1	3	1	1	0
4	0	1	2	1	0	3	1	1	14	2	5	3	1	5	1	2	1	0	5	5
4	0	1	0	0	2	3	0	4	1	0	4	2	1	11	1	16	4	2	6	1
4	0	4	15	1	3	3	1	3	1	0	1	1	0	3	0	0	0	1	1	1
0	2	0	1	3	0	3	2	0	4	2	1	5	1	2	5	0	2	15	1	2
1	5	2	0	2	2	2	4	1	0	3	1	2	3	0	4	0	2	14	1	6
16	17	0	0	6	1	2	1	1	5	2	0	5	2	2	12	2	2	16	2	3
1	3	3	0	3	3	0	2	3	2	14	2	1	6	4	3	15	3	4	16	2
3	0	4	3	2	2	1	5	2	1	0	0	0								

* * * *CONFIDENCE INTERVAL ANALYSIS * * * *
GARTNER LEE LIMITED
SIMULATION RUN NUMBER VV
11-21-1988
21:32:33
STATISTICS CALCULATED USING POPULATION MEAN (Mu)

DATASET c:\DESIGN\DATA00VV.PRN

MONTH 1

FR	N	UCL	MEAN	LCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
D	30	0.85	0.60	0.35	0.49	0.66	0.44	0.09	0.66	-0.63	0.00	1
T	13	1.20	0.77	0.34	0.87	0.72	0.51	-0.11	1.00	-0.53	28.21	2
W	4	1.31	0.50	0.00	1.62	0.51	0.26	0.00	-0.57	-1.85	-16.67	7

MONTH 2

FR	N	UCL	MEAN	LCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
D	30	5.03	2.90	0.77	4.25	5.71	32.62	0.44	2.10	2.61	0.00	1
T	13	8.70	4.38	0.06	8.64	7.15	51.06	0.55	1.79	0.73	51.19	2
W	4	4.22	1.00	0.00	6.45	2.03	4.11	0.00	-1.17	-1.56	-65.52	7

MONTH 3

FR	N	UCL	MEAN	LCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
D	30	2.67	1.60	0.53	2.15	2.88	8.31	0.13	4.11	17.82	0.00	1
T	13	1.76	1.08	0.40	1.36	1.13	1.27	-0.28	-0.65	-1.34	-32.69	2
W	4	16.53	5.00	0.00	23.16	7.20	53.06	-0.43	1.93	0.82	212.50	7

MONTH 4

FR	N	UCL	MEAN	LCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
D	30	35.40	19.27	3.14	32.26	43.31	1875.80	0.80	2.27	3.52	0.00	1
T	12	51.57	25.75	0.00	55.65	43.81	1919.45	0.43	2.31	3.68	23.27	2
W	5	23.76	0.80	0.00	45.96	18.48	341.58	-0.02	-1.00	-1.99	-95.85	7

MONTH 5

FR	N	UCL	MEAN	LCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
D	30	4.66	2.90	1.14	3.53	4.74	22.42	0.03	2.44	4.53	0.00	1
T	13	4.77	2.46	0.15	4.62	3.82	14.59	-0.08	2.30	4.85	-15.12	2
W	4	16.74	5.25	0.00	22.98	7.23	52.21	-0.37	1.84	0.63	81.03	7

MONTH 6

FR	N	UCL	MEAN	LCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
D	30	7.76	5.33	2.91	4.84	6.50	42.29	-0.24	1.26	-0.06	0.00	1
T	13	8.87	5.15	1.43	7.44	6.15	37.85	0.58	1.21	-0.12	-3.37	2
W	4	24.87	9.25	0.00	31.24	9.82	96.53	-0.86	1.19	-1.12	73.44	7

MONTH 7

FR	N	UCL	MEAN	LCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
D	30	2.77	1.80	0.83	1.95	2.61	6.83	-0.06	3.38	13.01	0.00	1
T	12	4.96	2.58	0.21	4.76	3.74	14.02	-0.29	2.87	6.41	43.52	2
W	5	3.94	1.80	0.00	4.28	1.72	2.96	-0.01	1.02	-0.35	0.00	7

MONTH 8

FR	N	UCL	MEAN	LCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
D	30	4.11	2.83	1.56	2.55	3.43	11.74	-0.11	2.23	5.52	0.00	1
T	13	6.74	3.92	1.11	5.64	4.66	21.72	0.49	2.14	2.67	38.46	2
W	4	5.01	1.00	0.00	8.02	2.52	6.36	-0.23	-1.04	-1.79	-64.71	7

MONTH 9

FR	N	UCL	MEAN	LCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
D	30	2.70	1.67	0.63	2.07	2.78	7.76	0.08	3.63	14.60	0.00	1
T	13	1.64	0.85	0.05	1.59	1.31	1.73	0.38	-0.83	-1.62	-49.23	2
W	4	[illegible]	[illegible]	0.00	4.71	1.48	2.19	-0.03	0.60	-1.05	5.00	7

MONTH 10

FR	N	UCL	MEAN	LCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
D	30	[illegible]	[illegible]	1.58	2.05	2.75	7.57	-0.21	2.99	10.95	0.00	1
T	13	3.09	[illegible]	1.37	1.71	1.42	2.01	-0.11	0.01	-0.59	-14.20	2
W	4	[illegible]	[illegible]	0.00	6.56	2.06	4.26	-0.94	-0.15	-1.77	-3.65	7

MONTH 11

FR	N	UCL	MEAN	LCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
D	30	[illegible]	[illegible]	2.83	4.20	5.64	31.80	0.17	1.25	[illegible]	0.00	1
T	12	[illegible]	[illegible]	1.57	7.19	5.66	32.03	-0.09	1.11	-0.37	4.73	2
W	5	13.17	5.00	0.00	16.34	6.57	43.20	-0.37	1.45	0.20	1.35	7

MONTH 12

FR	N	UCL	MEAN	LCL	RANGE	SD	VAR	AC	SKEW	EXCESS	DELTA	FREQ
D	30	5.32	3.83	2.35	2.97	3.99	15.94	-0.11	2.08	3.44	0.00	1
T	13	5.74	3.69	1.65	4.09	3.39	11.46	0.22	1.85	3.53	-3.68	2
W	4	15.56	5.50	0.00	20.12	6.33	40.03	-0.41	1.75	0.43	43.48	7

Computation time = 3.0 minutes.